Behavioral and Electrophysiological Measures of Speech-in-Noise Perception in Normal Hearing and Hearing Impaired Adults

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Tess Kathryn Koerner

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Yang Zhang, Ph.D., Peggy B. Nelson, Ph.D.

July 2017

Acknowledgements

I thank my primary advisor, Yang Zhang, for all of his time, effort, wisdom, and generous support over the years. I also thank my co-advisor, Peggy Nelson, for her continuous encouragement and guidance. I am exceedingly fortunate to have you both as mentors.

I offer sincere gratitude to my other committee members, Bert Schlauch and Andrew Oxenham, whose helpful insights and advice is greatly appreciated. I also thank Aparna Rao for her unwavering support.

A special thanks to Edward Carney for his substantial technical assistance and guidance in developing and implementing the last experiment of this dissertation and for endless programming and statistical support. Additional thanks to Boxiang Wang and Hui Zou for statistical support.

I would also like to recognize Zhang and Nelson lab members for their assistance and friendship: Sharon Miller, Yingjiu Nie, Luodi Yu, Chieh Kao, and Adam Svec. I would also like to express gratitude to the faculty and staff of Shevlin Hall, who have all helped me in too many ways to count over the years.

Finally, I would like to thank my parents, family, and friends for their endless encouragement, love, and support.

This dissertation project was made possible by funding from the Charles E. Speaks Graduate Fellowship, the Graduate Research Partnership Program (GRPP), and the Bryng Bryngelson Research Fund. Research projects were also supported in part by the Capita Foundation and Brain Imaging Research Project award to Yang Zhang. Participant payment for the first four studies was also partially supported by NIDCD R01-DC008306 to Peggy Nelson.

Abstract

Understanding speech in background noise is difficult for many individuals. Mechanisms responsible for variability in speech-in-noise performance across individuals are not well understood. Electrophysiological measures allow for an examination of the timing and strength of neural responses to speech along the auditory pathway and can be used to explore mechanisms underlying reduced speech perception in noise.

This dissertation used behavioral and electrophysiological measures to examine the effects of background noise on the neural coding of speech and to identify potential neural correlates of speech perception in individuals with and without hearing impairment. N1-P2, mismatch negativity (MMN), and P3 auditory event-related potentials (AERPs) and associated event-related cortical oscillations in various frequency bands of interest were collected in response to syllable-level stimuli in noise. Behavioral measures consisted of phoneme discrimination and sentence recognition in noise.

Results indicated that in addition to impacting averaged AERP responses, background noise disrupted cortical oscillatory rhythms in response to speech in frequency bands of interest across participants. Results also showed that the effects of background noise and hearing impairment on the neural coding of speech are different at different levels of cortical processing. This work revealed that AERPs and associated cortical oscillations represent potential neural correlates of speech perception in noise in individuals with and without hearing impairment. These findings have potential theoretical and practical implications regarding the use of electrophysiological measures for the assessment and rehabilitation of communication difficulties in background noise.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

I. Overview

Listening in background noise presents a communication challenge for many individuals with or without hearing impairment (HI). The standard audiological evaluation assesses hearing thresholds using pure-tone stimuli in quiet. However, it has been established that this pure-tone assessment may not be able to accurately assess or predict the ability to perceive speech across individuals with normal hearing (NH) or those with HI, especially in complex listening environments (Dubno, Dirks, & Morgan, 1984; Hargus & Gordon-Salant, 1995; Jin & Nelson, 2006; Killion & Niquette, 2000; Kujawa & Liberman, 2009; Souza, Boike, Witherell, & Tremblay, 2007; Vermiglio, Soli, Freed, & Fisher, 2012). Although word-level stimuli are often used to assess suprathreshold speech perception during standard audiological evaluations, these measures are typically only assessed in quiet listening conditions and are unable to accurately predict performance in background noise (Wilson, 2011). The development of more recent speech-in-noise tests, such as the Hearing-in-Noise-Test (Nilsson, Soli, & Sullivan, 1994) or the Words-in-Noise-Test (Wilson, 2003), allow for a better indication of a client's abilities to perceive speech in noise. However, these tests are still unable to predict performance in more "real world" listening environments and do not allow for a systematic examination of peripheral and central processing mechanisms that may impact performance.

While the exact cause of substantial variability in speech-in-noise perception across individuals is currently unknown, various sensory and cognitive factors likely

contribute to the ability to communicate in noise. It has been suggested that in addition to reduced audibility, impaired suprathreshold spectrotemporal processing skills or reductions in cognitive ability, such as impaired auditory attention, memory, or processing speed, may contribute to reduced performance in noise in individuals with HI (Bacon, Opie, & Montoya, 1998; Gordon-Salant & Fitzgibbons, 1997; Peters, Moore, & Baer, 1998; Shinn-Cunningham & Best, 2008). These potential peripheral and central processing deficits may impact the ability to process acoustic cues that are important for speech perception. While it is well established that acoustic cues within consonants and vowels are highly overlapped in connected speech, it has been shown that the relative contributions of these segments to speech perception may depend on listening context (Cole, Yan, Mak, Fanty, & Bailey, 1996; Fogerty & Humes, 2012; Fogerty & Kewley-Port, 2009; Kewley-Port, Burkle, & Lee, 2007). For instance, research has shown that vowels carry the most important acoustic information for facilitating sentence-level perception compared to consonant segments, while the relative contribution of consonants and vowels appear to be the same for the perception of isolated words (Fogerty & Humes, 2010). However, how background noise alters the neural coding of consonants and vowels at different stages of cortical processing, and how the effects of background noise impact speech perception in noise at the phonemic and sentential levels in individuals with NH or HI is not well understood. This information is integral to better understanding the causes of variability and decreased performance in speech perception in noise across individuals.

These issues highlight the importance of using other measures to assess auditory processing and speech perception. Non-invasive electrophysiological measures can be used to examine the timing and magnitude of neural responses to speech in noise along the auditory pathway (Luck, 2014). When paired with behavioral tasks, these measures can provide complementary information about the neural mechanisms underlying speech perception in noise. Since they are objective measures of neural processing, electrophysiological measures may be useful clinical tools for predicting a patient's ability to perceive speech. However, in order to use these measures clinically, we need a better understanding of how background noise and hearing loss impact the neural coding of speech during different stages of cortical processing and how variability in the neural coding of speech impacts behavioral performance.

This dissertation project used behavioral and electrophysiological measures to examine the neural coding and perception of speech in noise in participants with and without HI. This work aimed to better understand how HI and background noise impact the neural coding of consonant and vowel stimuli and was designed to determine whether variability in the neural responses to these stimuli are predictive of sentence-level perception in noise across participants. The behavioral tests used in this study measure percent correct sentence recognition, percent correct phoneme detection accuracy or sensitivity, as well as phoneme detection reaction time. The following section provides background information on electrophysiological measures and recording paradigms.

II. Electrophysiology Background

A. Auditory Event-Related Potentials

Event-related potentials are thought to be a sum of the voltages of synchronous fixed latency and fixed polarity post-synaptic electrical activity that is time-locked to repeated stimuli (Luck, 2014). These potentials represent neural activity in localized brain regions that are responsible for processing different aspects of information related to perceptual stimuli or tasks. This definition represents the classic "evoked" or "additive model," in which ERP components are superimposed on, and independent from, random background EEG activity. Since these synchronous potentials are fixed in latency and polarity, the evoked response becomes larger and ongoing EEG "noise" becomes smaller when trials are averaged together. This averaging results in an ERP waveform that consists of several positive and negative voltage deflections. These ERP waveform peaks are analyzed in terms of their latency and amplitude. Latency represents neural processing time and is measured in milliseconds (ms) while amplitude is a measure of the magnitude of neural activation to the presented stimulus and is measured in microvolts (μV). Event-related potentials have precise timing, which makes them desirable for studying time-varying signals such as speech. However, ERP averaging requires hundreds of stimulus trials, and therefore these recording paradigms typically use short duration pure-tones, complex-tones, or speech syllables instead of longer duration speech utterances such as sentences. Late cortical potentials consist of neural activity that occurs more than 80 ms post-stimulus onset and are characterized by a series of ERP peaks that are often used to study the neural mechanisms underlying speech perception. These include the N1-P2 complex, the mismatch negativity (MMN) response, and the P3 response (Luck, 2014; Martin, Tremblay, & Korczak, 2008).

N1-P2 Complex

The N1-P2 complex is an obligatory AERP that signals the neural encoding of audible acoustic information in the auditory cortex. This sensory potential is typically recorded during a passive listening task using repeated auditory stimuli. The N1 is seen as a negative peak about 100 ms post-stimulus onset and the P2 is seen as a positive peak at about 175 ms post-stimulus onset. These ERPs are typically recorded at mid-central electrodes, such as electrode Cz, as neural generators are thought to include the primary and secondary auditory cortices (Näätänen & Picton, 1987).

The Mismatch Negativity (MMN) Response

The MMN is also an obligatory AERP recorded during a passive listening task, however, it is elicited using an oddball paradigm with a common, repeating "standard" stimulus and less frequently occurring "deviant" stimuli. It is the response to the more infrequent deviant stimulus which violates the repetitive pattern of the standard stimuli from the participants' short-term memory trace. The resulting MMN waveform is a difference wave, which is found by subtracting the response to the "standard" stimuli from that to the "deviant" stimulus. The MMN is thought to signal pre-attentive auditory discrimination and is typically recorded in the absence of any overt behavioral response or attention. For example, participants are often asked to ignore the auditory stimulus presentations while focusing on a muted, subtitled movie. The MMN peak is seen as a negative deflection at about 100-300 ms after the onset of a discriminable stimulus change. It is typically recorded at frontocentral electrode sites, such as Fz or Cz. Researchers have detected at least two neural generators that are responsible for the

MMN: a bilateral supratemporal process in the auditory cortex that is responsible for stimuli change detection and a frontal process that is responsible for an attention switch toward the deviant stimuli (Kraus, McGee, Carrell, & Sharma, 1995; Näätänen, 1995; Näätänen, Paavilainen, Rinne, & Alho, 2007; Stapells, 2002).

The P3 Response

The P3 is an active, cognitive ERP that reflects the ability to discriminate stimuli presented in an oddball paradigm. This response is recorded while participants are paying attention and actively responding to the stimuli being presented. For example, participants may be asked to push a button whenever a deviant stimulus is heard. The P3 peak is typically seen as a positive deflection peak that occurs approximately 300 ms after stimulus change onset. It is recorded at paritocentral sites, such as Pz. It has been suggested that the P3 is generated by multiple cortical regions that are activated based on the specific task at hand, but it is still not well understood. Some have suggested that it reflects "closure" in the decision making process after the participant responds (Picton, 1992; Stapells, 2002).

B. Cortical Oscillations

Averaged ERP waveforms are by far the most popular electrophysiological method for examining the neural coding of speech. However, there is an ongoing debate about the neural underpinnings of ERP generation. An alternate view, known as the "phase-resetting hypothesis" or "oscillation theory," suggests that oscillations within the ongoing EEG signal are reset by the onset of a stimulus, which causes neurons to begin firing synchronously in the excitatory phase. The ongoing EEG signal can be broken

down into separate oscillatory frequency bands using a Fourier transform and can be extracted and analyzed using time-frequency analysis techniques. In this view, ERP components are generated by averaging the phase-reset EEG signals across frequency bands without any additive evoked neural response. These EEG oscillations are thought to control the timing of synchronous neural firing in the excitatory phase in order to enable perceptual, behavioral, and cognitive processes within and across cortical networks. Two common techniques used to examine event-related cortical oscillations in different frequency bands at time points of interest are measures of trial-by-trial phase locking (also known as neural synchrony or inter-trial phase coherence (ITPC)) and spectral power analysis. Variations in the phase and amplitude of oscillatory activity likely reflect neural excitability and synchronous neural firing elicited in response to a perceptual task (for a review, see Klimesch, Sauseng, Hanslmayr, Gruber, & Freunberger, 2007; Sauseng et al., 2007).

A drawback of using traditional ERP analysis techniques is that any trial-by-trial oscillatory activity that is not phase-locked to the stimulus is removed during the averaging process. For instance, if the power of oscillatory activity is strong, but trial-by-trial neural synchrony is disrupted, the averaged ERP response may reveal falsely prolonged peak latencies or reduced peak amplitudes or may even result in little to no measurable response. In other words, the averaged AERP waveform does not allow for an examination of variability in the neural response across stimulus trials, even though these underlying oscillatory rhythms may be influencing sensory and cognitive processes related to a specific experimental stimulus, task, and behavioral response. Instead, time-

frequency analysis techniques allow researchers to consider more information from the ongoing EEG signal and approach human neurophysiological research using a new conceptual framework. This more refined analysis can provide information about what frequency bands are influencing perception and how they contribute to the neural generation of ERP components.

Cortical EEG oscillations have been functionally associated with various sensory and cognitive processes. For instance, it has been shown that theta oscillations (4-8 Hz) are related to memory encoding, retrieval, and maintenance processes (Ward, 2003). Thus, it is not surprising that these oscillatory rhythms have been shown to contribute to the neural generation of the auditory MMN response (Choi et al., 2013; Fuentemilla, Marco-Pallarés, Münte, & Grau, 2007; Ko et al., 2012). Similarly, research has suggested that neural oscillations in delta (0.5-4 Hz) and theta (4-8 Hz) frequency bands correlate with decision making and focused attention, respectively, and are dominant components in driving the generation of the P3 response (Basar-Eroglu, Basar, Demiralp, & Schurmann, 1992). Even spontaneous EEG activity has been shown to be related to behavioral perception. Studies that have explored resting state or pre-stimulus baseline EEG activity have linked spontaneous EEG rhythms to cortical excitation as well as perceptual processes underlying behavioral responses (Ergenoglu et al., 2004; Hanslmayr et al., 2007; Romei et al., 2008; Sauseng, Klimesch, Gerloff, & Hummel, 2009). These findings highlight the importance of using time-frequency analysis techniques in addition to traditional ERP waveform averaging techniques, as examining cortical oscillatory

activity can provide complementary information about sensory and cognitive processes underlying speech perception.

C. Experimental Protocols and the Double Oddball Paradigm

N1-P2 responses are typically recorded using different ERP experimental protocols and recording paradigms than MMN and P3 responses. The N1-P2 complex can be elicited using conventional block presentations, where hundreds of repetitions of the same auditory stimulus are presented to the participant in a single block (Billings, Tremblay, Souza, & Binns, 2007; Tremblay, Piskosz, & Souza, 2003). This paradigm has since been modified into an improved alternating short block design, which requires less recording time and can reduce neural response adaptation or habituation that often occurs in response to long blocks of repeated, identical stimuli (Zhang et al., 2011). This paradigm contains short blocks of repeated stimulus presentations with equal stimulus ratios and is useful for comparing the neural processing of different acoustic stimuli, including phonemic speech contrasts. It has been shown that N1-P2 responses recorded with this alternating short block paradigm are sensitive to variations in the acoustic features of different speech stimuli that are presented in alternating blocks as well as the effects of different listening conditions on the neural coding of these stimuli (Miller & Zhang, 2014; Zhang et al., 2011). However, while the single block and alternating short block designs allow for an examination of the cortical encoding of acoustic features of different stimuli, they are not appropriate for measuring sensory discrimination abilities.

The MMN and P3 AERPs, which signal pre-attentive and active change detection processes, respectively, are typically recorded using an oddball paradigm. The

conventional oddball paradigm consists of a repeating train of a standard stimulus that is occasionally violated by a less frequently occurring deviant stimulus that has a probability of occurrence of approximately 10% (Näätänen et al., 2007). There are several variants of this conventional oddball paradigm that have also been used to study the sensory discrimination of speech contrasts. For instance, double-oddball or multi-feature oddball paradigms can include two or more deviant stimuli with equal probability of occurrence that vary from the standard stimulus in a single acoustic feature (Näätänen, Pakarinen, Rinne, & Takegata, 2004; Pakarinen et al., 2013; Pakarinen, Takegata, Rinne, Huotilainen, & Näätänen, 2007). These paradigms have several advantages over the traditional oddball paradigm, including the fact that they are less time-consuming and allow for an examination of neural responses to several juxtaposed deviant speech contrasts or speech categories that are in direct competition with one another for neural resources. For example, a double-feature oddball paradigm essentially allows for a within-subject control condition while examining differences in the cortical encoding of two different speech contrasts or while exploring the effect of various experimental factors on the neural processing of different acoustic cues.

III. Overview of Planned Studies

Non-invasive electrophysiological measures can provide information about speech processing at various stages along the auditory pathway. Ongoing evidence suggests that event-related cortical oscillations likely play a large role in controlling the timing of neural firing during important sensory and cognitive events and at least partially contribute to the neural generation of ERPs. Examining how cortical

oscillations in various frequency bands are impacted by background noise and whether they are linked to the ability to perceive speech in noise will shed light on their functional significance, will further our understanding about the neural generation of ERPs, and will establish whether EEG oscillations can be used as an index of speech perception in noise. The five experiments included in this dissertation aim to further understand speech processing mechanisms affected by the presence of background noise in individuals with and without hearing loss as well as to determine whether AERPs and their associated event-related cortical oscillations are able to predict speech perception across individuals. The specific research questions and expected outcomes for each study are outlined below:

A. Study 1

Research questions:

1. How does the introduction of background noise impact ITPC in delta, theta, and alpha bands at time points corresponding to N1 and P2 responses in normal-hearing participants?
2. Do noise-induced changes in ITPC across frequency bands account for variability in N1-P2 latencies and amplitudes in response to speech?

Expected Outcomes:

1. The introduction of background noise will increase N1-P2 latencies, decrease N1-P2 amplitudes, and decrease ITPC across frequency bands.
2. Variations in trial-by-trial neural synchrony will predict noise-induced increases in N1-P2 latencies and decreases in N1-P2 amplitudes across participants with normal hearing.

B. Study 2

Research questions:

1. Does background noise have a differential effect on MMN and behavioral responses to a consonant and vowel change?
2. Are noise-induced changes in the MMN reflected in variability in spectral power in the theta band?
3. Are objective MMN latency, amplitude, and theta power measures recorded in response to a consonant and vowel change able to predict phoneme- and sentence-level perception in background noise across participants with normal hearing?

Expected Outcomes:

1. Background noise would increase MMN latency, decrease MMN amplitude, and decrease spectral power in the theta band across speech stimuli, but that that there would be larger noise-induced effects on the neural coding and perception of the consonant change compared to the vowel change.
2. Noise-induced changes in theta power would be reflected in variability in averaged MMN latencies and amplitudes.
3. Objective MMN responses would represent potential neural correlates of speech perception in noise across participants.

C. Study 3

Research question:

1. Are mixed-effects regression models more appropriate than Pearson correlation analysis for modeling multivariate neurophysiological in the examination of brain-behavior relationships?

Expected Outcome:

1. A comparison of statistical methods would show that mixed-effects models are more capable of handling repeated measures data that contain built-in within-subject contrasts than traditional Pearson correlations.

D. Study 4

Research questions:

1. How does background noise impact P3 latency, amplitude, and trial-by-trial neural synchrony and power in delta, theta, and alpha bands in response to a consonant and vowel change in normal-hearing participants?
2. Does attention modulate the effect of background noise on the neural coding of consonant and vowel stimuli?
3. Are noise-induced changes in the P3 AERP and event-related cortical oscillations in delta, theta, and alpha frequency bands predictive of changes in phoneme- and sentence-level perception in noise across normal-hearing participants?

Expected Outcomes:

1. Background noise would increase P3 latency, decrease P3 amplitude, and decrease trial-by-trial phase locking and spectral power across frequency bands in response to both the consonant and vowel change.

2. Focused attention would have a differential effect on the neural coding of the consonant and vowel change.
3. The P3 AERP and its associated event-related cortical oscillations recorded in response to a consonant and vowel change in noise would be predictive of behavioral performance on speech-in-noise tests across normal-hearing participants.

E. Study 5

Research questions:

1. Does hearing impairment have a differential effect on N1-P2 and MMN responses to speech in background noise?
2. Can N1-P2 and MMN AERPs and their associated event-related cortical oscillations recorded in response to speech in noise predict the effects of hearing impairment on speech-in-noise perception?

Expected Outcomes:

1. Hearing sensitivity would have a larger impact on later cortical potentials that reflect perceptual and cognitive speech discrimination processes, as measured by the MMN, compared to earlier N1-P2 responses that reflect the sensory encoding of acoustic information.
2. Hearing-related changes in both the N1-P2 complex and the MMN response would be predictive of the effects of hearing impairment on speech perception in background noise.

Chapter 2: Examination of the Effects of Noise on Event-Related Cortical Oscillations and AERPs

Sections I-V are reprinted from:
Koerner, T.K. and Zhang, Y. (2015). Effects of Background Noise on Inter-Trial Phase Coherence and N1-P2 Responses to Speech Stimuli. *Hearing Research,* 328, 113-119.

I. Introduction

Perceiving speech in noise can be a challenge for people with or without hearing impairment (HI). Emerging research on cortical auditory event-related potentials (AERPs) has established that the presence of background noise affects the neural processing of speech (Bennett, Billings, Molis, & Leek, 2012; Billings, Bennett, Molis, & Leek, 2011; Kaplan-Neeman, Kishon-Rabin, Henkin, & Muchnik, 2006; Martin, Sigal, Kurtzberg, & Stapells, 1997; Martin & Stapells, 2005; Martin, Kurtzberg, & Stapells, 1999; Whiting, Martin, & Stapells, 1998). Cortical AERPs provide an objective time-domain measure of the timing and strength of cortical responses to auditory stimuli. The averaged latencies and amplitudes of these AERPs are influenced by several physiological factors, including within- and across-trial neural synchrony (Eggermont, 2007).

In an attempt to probe the neural synchrony that underlies the AERP, a number of studies have measured the influence of inter-trial phase coherence (ITPC or ITC; also referred to as phase locking value (PLV)) in certain characteristic frequency bands on the AERP (Başar, Demiralp, Schürmann, Başar-Eroglu, & Ademoglu, 1999; Demiralp et al., 1999; Luck, 2014). For instance, studies on normal hearing children and adults suggest that inter-trial synchronization of cortical oscillations in lower frequency bands such as theta and alpha plays an important role in determining the amplitude of AERP responses

(Bishop, Anderson, Reid, & Fox, 2011; Edwards et al., 2009). A recent study further demonstrated the potential clinical utility of the ITPC approach in assessing cortical neural synchrony in relation to severity of sensorineural hearing loss, listening experience, and aural rehabilitation improvement in young children (Nash-Kille & Sharma, 2014).

In general, cortical oscillations are thought to reflect fluctuations in neural excitability, and as such, are thought to reflect information exchange between and within cortical networks (Buzsáki & Draguhn, 2004; Herrmann & Knight, 2001; Pantev, Elbert, & Lütkenhöner, 1994; Schroeder & Lakatos, 2009). In the domain of auditory neuroscience in particular, time-frequency analysis has shown significant enhancement/reduction effects in cortical oscillatory activities that are associated with corresponding changes in AERP components as a function of stimulus/task manipulation (Başar et al., 1999; Başar, Başar-Eroglu, Karakaş, & Schürmann, 2000; Pantev, Elbert, & Lütkenhöner, 1994; Zhang et al., 2011).

While researchers have no disputes about the prevalence of various oscillatory rhythms in EEG data, there are ongoing debates about their functional significance in the generation of the ERPs (Bastiaansen, Mazaheri, & Jensen, 2012; Klimesch, Sauseng, Hanslmayr, Gruber, & Freunberger, 2007; Makeig, Debener, Onton, & Delorme, 2004). The classical view is referred to as an "additive "model, which assumes neural oscillatory activities that can be observed prior to stimulus presentation have random phase relationship whereas the evoked signal that is responsible for stimulus coding/processing would remain relatively stable in each trial. The averaging process for ERPs would thus

cancel out the ongoing oscillatory activities that are not time-locked to the stimuli. An alternative view, termed the "phase resetting" model, however, shows that the ERP components can arise from resetting the phase of the ongoing oscillations at the onset of acoustic stimulation in each trial. In other words, the generation of the AERP responses does not require any power gain from the relatively stationary "evoked" signal for stimulus coding in each trial. In mathematical simulation, it is compelling that the two models can be equally valid to represent physiologically independent neural mechanisms for the origin of ERP responses. Since both "evoked" and "phase reset" signals are phasic signals and may share physical characteristics, the ITPC measure is likely to capture phase-locked oscillatory rhythms in both instances without being able to dissociate the potential contribution of each mechanism. Therefore, one argument in favor of the phase resetting model for the origin of AERPs is to demonstrate that the spectral power of the neural oscillations involves no noticeable change in the pre- and post- stimulus time windows of the AERP (Fuentemilla, Marco-Pallarés, & Grau, 2006; also see reviews by Bastiaansen, Mazaheri, & Jenson, 2012; Klimesch et al., 2007).

The present study aimed to examine the effects of background noise on speech processing in terms of ITPC and the evoked auditory N1-P2 responses. The auditory N1-P2 responses are considered obligatory AERP components that signal the neural processing of important acoustic features in speech sounds (Eggermont, 2007). It has been consistently demonstrated that the auditory N1 and P2 latencies tend to increase and their amplitudes tend to decrease in adverse listening environments. Such changes in AERP can occur when high- or low-pass filters alter the spectral content of the masker

(Martin et al., 1997; Martin & Stapells, 2005; Martin et al., 1999), when the masking noise has a higher intensity level giving rise to poorer signal-to-noise ratios (SNR) (Kaplan-Neeman et al., 2006; Whiting et al., 1998), or when it includes informational masking effects (Billings et al., 2011). While these results reveal the effects of background noise on AERP latency and amplitude measures, an important question remains unaddressed, i.e., whether the noise-induced changes in the N1-P2 responses are related to possible changes in ITPC in response to the speech stimuli. Previous research has established that neural oscillations in the delta (0.5-4 Hz), theta (4-8 Hz), and alpha (8-12 Hz) frequency bands contribute to sensory and cognitive processes related to audition as reflected in the obligatory N1-P2 responses during the neural encoding of speech sounds (Başar et al., 1999; Başar et al., 2000; Pantev et al., 1994; Zhang et al., 2011). In particular, Edwards et al. (2009) demonstrated the existence of significant correlations between ITPCs and ERPs for the theta and alpha bands but not for the higher frequency bands such as gamma (> 30 Hz). However, it remains unclear how neural synchrony as assessed by ITPC in these different frequency bands are adversely affected by noise in relation to the latency and amplitude measures of the obligatory AERP responses.

We designed the study with the purpose of replicating previous findings of modulatory effects on N1 and P2 responses by background noise and investigating further whether changes in ITPC in quiet and noise (an in-depth complimentary measure looking at trial-by-trial neural synchrony) could account for the variations in N1-P2 latencies and amplitudes for speech processing in the two listening conditions. We

hypothesized that the presence of background noise would reduce inter-trial phase locking and that the reduction in phase locking would be correlated with changes in N1 and P2 latency and amplitude measures. More specifically, we predicted that disrupting neural synchrony with the introduction of background noise would introduce a neural timing delay and a reduction in the magnitude of neural responses in response to speech. The results have important implications regarding the interpretation of averaged AERP waveforms as well as the modulatory effects of background noise on the neural representation of speech. Our approach can also add valuable evidence in the assessment of the additive vs. phase resetting models. For instance, the phase resetting model would receive additional support if the trial-by-trial averaged power spectrum did not differ between the quiet and noise listening conditions.

II. Methods

A. Subjects

The participants in the study were 11 individuals (mean age = 23.6 years, age range = 20 – 32 years, 5 males, 6 females) with normal hearing sensitivity (as shown in standard audiologic assessment with hearing thresholds < 25 dB HL for pure tones from .25 to 8 kHz) and no history for speech, language, or cognitive disorders. All participants were right-handed and were native speakers of American English. The research protocol was approved by the Human Research Protection Program at the University of Minnesota. Informed consent was obtained from all subjects, who were paid for their participation.

B. Stimuli

The stimuli were taken from a previous study on speech perception in noise (Koerner, Zhang, Nelson, 2013). The speech syllable /bu/ was synthesized with the HLsyn software program (Sensimetrics Corporation, USA) at a sampling rate of 10 kHz. The fundamental frequency was set at 100 Hz, and the stimulus duration was 170 ms. The syllable consisted of four formants with F4 fixed at 3300 Hz for the entire syllable. The onset frequencies for F1, F2, and F3 were set at 230 Hz, 900 Hz, and 2480 Hz, and the steady F1, F2, and F3 center frequencies for the vowel portion (50 – 170 ms) were respectively 320 Hz, 860 Hz, and 2620 Hz. Formant transitions in the first 50 ms of the syllable were automatically generated by the HLsyn program. The four-talker speech babble masker was adopted from the Quick Speech In Noise Test (Quick-SIN) (Etymotic, 2001). Both the speech syllable and noise stimuli were resampled at 44.1 kHz, and normalized to create a -3 dB SNR using Sony SoundForge 9.0 (Sony Creative Software, USA) for the listening in noise condition.

C. Procedure

All testing was performed in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems). The reported data were taken from one passive listening session in a larger-scale study to examine neural correlates of speech intelligibility performance (Koerner, Zhang, & Nelson, 2013). During the passive listening session, the participants were instructed to minimize extraneous eye or muscle movements and stay awake and relaxed while watching a movie of their own choice with English subtitles and ignoring the auditory stimuli played through headphones. The stimuli were presented using the EEVoke software (ANT Inc., Netherlands) via bilateral Etymotic ER-2 insert

earphones. Participants were presented with two different listening conditions: speech in quiet and speech in noise with the four-talker speech babble. The speech stimulus was presented at a level of 60 dB SL relative to individual subjects' hearing threshold at 1 kHz with mean RMS level matched to that of the /bu/ syllable. The babble noise was added as continuous background at a -3 dB SNR for the speech in noise condition. There were 120 trials presented in 4 blocks for each listening condition. The interstimulus interval (ISI) between any two consecutive /bu/ stimuli in each block used a random number between 600 ms and 700 ms. The presentation order for the two conditions was counterbalanced among the subjects.

D. Data Analysis

Continuous electrophysiological data were recorded with the Advanced Neuro Technology EEG System (Advanced Source Analysis version 4.7) and a 64-channel Ag AgCl electrode WaveGuard cap with a REFA-72 amplifier (TMS International BV) (bandwidth = 0.016-200 Hz, sampling rate = 512 Hz). The average impedance of electrodes was below 5 kOhms. The same recording was used in previous ERP studies (Rao, Zhang, & Miller, 2010; & Miller & Zhang, 2014). ERP waveform analysis was completed offline in BESA (Version 6.0, MEGIS Software GmbH, Germany) and MATLAB (Version 8.0). The EEG data were bandpassed at 0.5-40 Hz. Automatic artifact rejection criteria were set at ± 50 μV. The ERP epoch length consisted of a 100 ms pre-stimulus baseline interval and a 700 ms post-stimulus interval. N1-P2 responses were analyzed with an averaged mastoid reference at the Cz electrode that showed the largest N1 and P2 responses. Based on grand mean ERP waveforms and scalp potential

topography, search windows relative to the pre-stimulus interval were determined for N1 and P2 respectively at 80-160 ms and 120-200 ms for the quiet condition. The search windows for N1 and P2 were respectively at 100-200 ms and 140-300 ms for the noise condition. Averaged peak amplitudes were quantified with a 10 ms window centered at the peaks (Rao, Zhang, & Miller, 2010).

Neural synchrony in delta (0.5-4 Hz), theta (4-8 Hz), and alpha (8-12 Hz) frequency bands was calculated with the ITPC measure in the open source EEGLAB software (Delorme & Makeig, 2004): $ITPC_{(t,f)} = \frac{1}{n}\sum_{k=1}^{n} \frac{F_k(t,f)}{|F_k(t,f)|}$, where F stands for the Fourier transform, t is time, f is frequency, k is the trial number, and | | is the complex norm. The ITPC represents an estimate of mean normalized phase across the EEG trials as a function of the time point and frequency in the epoch time series. The normalized phase measure is obtained from the complex output of the frequency transformation by dividing by its complex norm for each trial and averaged across trials. The ITC values at a given latency can range from 0 (indicating absence of synchronization or phase reproducibility) to 1 (indicating perfect synchronization across trials). We used a modified short-term Fourier Transform (STFT with Hanning window tapering as implemented in EEGLAB), which has been shown to have relatively good resolution for low frequency activities. The frequency range was 0.5 to 40 Hz, and the step interval was 0.5 Hz. In the conventional STFT method, a fixed time window is applied to all frequencies. The uniformity of the time window is thus a limitation of this approach because optimal characterization of changes in higher frequency signals would require shorter time windows than those needed to characterize lower frequency signals. The

modified STFT method in EEGLAB uses overlapping sliding windows that are adaptive to the target frequency bins (i.e., the time window decreases linearly as frequency increases). Zero-padding is applied for short epochs without sufficient number of sample points for Fourier transform. For each frequency band in our calculation, the ITPC/PLV data were averaged across the frequencies within its range for further processing. The peak phase locking values corresponding to the N1 and P2 components in their respective windows were identified for each frequency band on an individual basis for each listening condition. The averaged phase locking values across the time samples in the baseline were also computed for comparison.

All statistical analyses were completed in R (R Core Team, 2014). A repeated measure analysis of variance (ANOVA), with $\alpha = 0.05$, was conducted to examine the statistical significance of listening condition (quiet vs. noise) on N1-P2 latencies and amplitudes recorded at electrode Cz. The repeated-measures ANOVA was also applied in evaluating the effect of listening condition on inter-trial phase locking in the three frequency bands. For each frequency band of cortical oscillations, Pearson's correlation analyses were performed on the amount of neural phase locking and N1-P2 amplitude/latency measures pooled across the two listening conditions.

III. Results

Overall, the presence of continuous background babble speech noise increased N1-P2 latencies and decreased N1-P2 amplitudes (Figs. 1 and 2). The time-frequency analysis revealed a consistent noise-induced reduction of trial-by-trial neural phase locking in all the delta, theta, and alpha frequency bands (Fig. 3). Correlational analyses

revealed that reductions in neural phase locking could predict changes in N1-P2 responses to speech in noise (Fig. 4).

A. Amplitude and Latency Measures for the averaged N1 and P2 responses

Table 1 shows the means and standard errors (SE) for N1 and P2 AERP latency and amplitude measures in response to the CV syllable /bu/. Grand mean ERP waveforms, scalp potential topography maps, and global field power (GFP) data for the quiet and noise listening conditions are shown in Figure 1. Given that there was large variability in the P2 latency values across our subjects for the noise condition (See Table 1, Fig. 2 and Fig. 4), it is not surprising that the P2 response in the grand mean at the Cz electrode appeared to be completely eliminated in the noise condition. Nevertheless, the GFP plot appeared to capture the noise-induced morphological changes to the N1-P2 response very well.

Repeated measures analysis of variance (ANOVA) revealed significant effects of background noise in the N1 latency ($F(1,10) = 53.71$, $p < 0.001$), N1 amplitude ($F(1,10) = 13.85$, $p < 0.01$), and P2 latency ($F(1,10) = 22.27$, $p < 0.001$). The effect of background noise on P2 amplitude showed a similar trend of reduction, but it did not reach statistical significance ($F(1,10) = 2.86$, $p = 0.122$). Compared with the quiet listening condition, N1 and P2 latencies for the /bu/ syllable in the noise were significantly delayed coupled with a significant reduction in the N1 amplitude.

Table 1. N1 and P2 mean latency (ms) and amplitude (μV) values (standard error) in response to the CV syllable /bu/ at electrode Cz in quiet and in noise.

	N1				P2			
	Quiet		Noise		Quiet		Noise	
Stimulus	Latency	Amplitude	Latency	Amplitude	Latency	Amplitude	Latency	Amplitude
/bu/	105.21 (3.46)	-2.91 (0.42)	165.87 (7.8)	-1.48 (0.21)	172.2 (3.77)	1.56 (0.4)	248.49 (15.43)	0.84 (0.23)

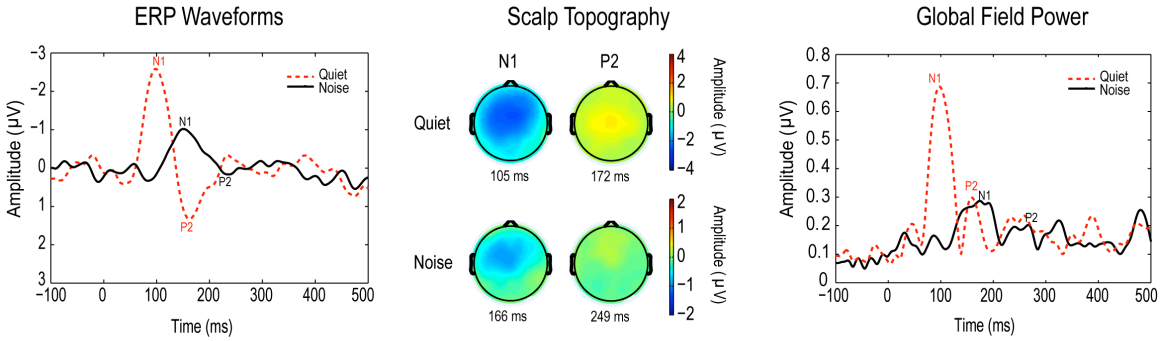

Figure 1. Grand mean AERP waveforms, scalp topography maps, and global field power plots averaged across participants in response to the CV syllable /bu/ in quiet (red dashed) and in noise (black) at electrode Cz. (See AERP waveforms from individual subjects in Figure 2).

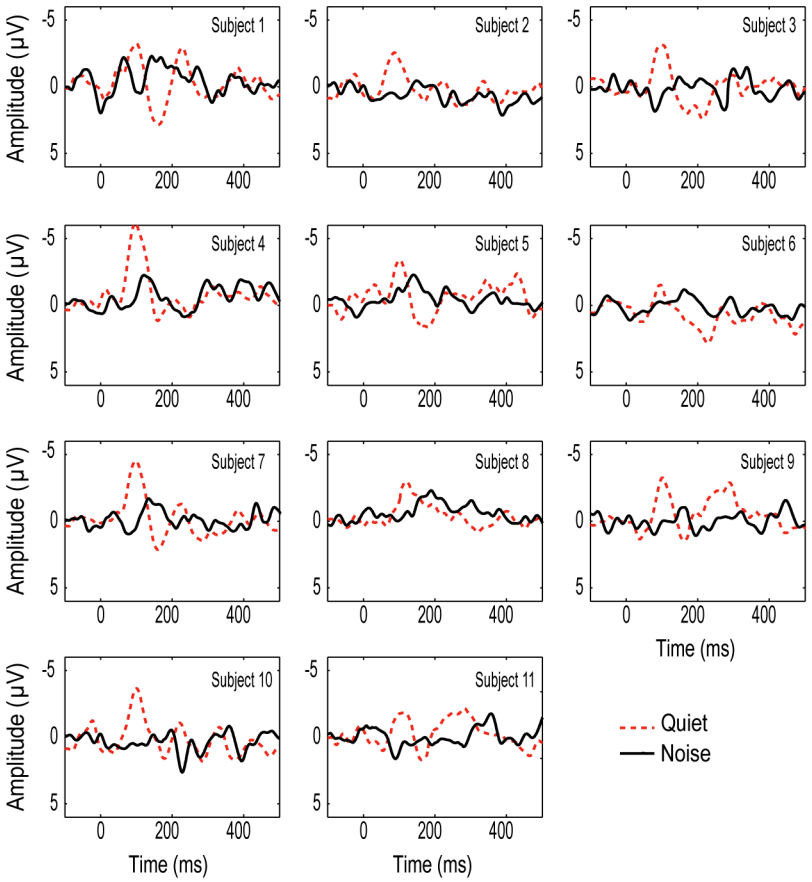

Figure 2. Individual AERP waveforms for all 11 participants in response to the CV syllable /bu/ in quiet (red dashed) and in noise (black) at electrode Cz.

B. Inter-trial Coherence in Phase Locking

Table 2 shows the trial-by-trial phase locking values for delta, theta, and alpha bands in response to the CV syllable /bu/ in quiet and in noise (See Fig. 3 for the ITPC/PLV plots and power density spectrum plots). Corresponding to the phase locking reduction for N1-P2, AERP power spectrum also showed decreased energy in the delta, theta, and alpha range in the noise condition relative to the quiet condition. However, the averaged trial-by-trial power spectrum data did not shown much difference between the two listening conditions.

Repeated measure ANOVA revealed significant noise-induced reduction in phase locking for N1 in the delta ($F(1,10) = 20.68$, $p < 0.01$), theta ($F(1,10) = 18.51$, $p < 0.01$), and alpha ($F(1,10) = 23.45$, $p < 0.001$) frequency bands. A similar significant reduction effect was found for P2 in the delta ($F(1,10) = 13.27$, $p < 0.01$), theta ($F(1,10) = 14.86$, $p < 0.01$), and alpha ($F(1,10) = 14.57$, $p < 0.01$) frequency bands. In contrast, the baseline phase locking data at the group level did not seem to be affected by the presence or absence of background speech-babble noise despite the existence of small changes at the individual subject level.

Table 2. Phase locking values for peak N1 and P2 in quiet and in noise and average baseline in quiet and in noise in response to the CV syllable /bu/ at electrode Cz.

Frequency Band	Baseline		N1		P2	
	Quiet	Noise	Quiet	Noise	Quiet	Noise
Delta	0.07 (0.01)	0.09 (0.01)	0.49 (0.03)	0.37 (0.01)	0.46 (0.03)	0.36 (0.01)
Theta	0.07 (0.01)	0.09 (0.01)	0.46 (0.03)	0.36 (0.02)	0.45 (0.03)	0.34 (0.01)
Alpha	0.07 (0.01)	0.09 (0.01)	0.47 (0.03)	0.36 (0.01)	0.45 (0.03)	0.34 (0.01)

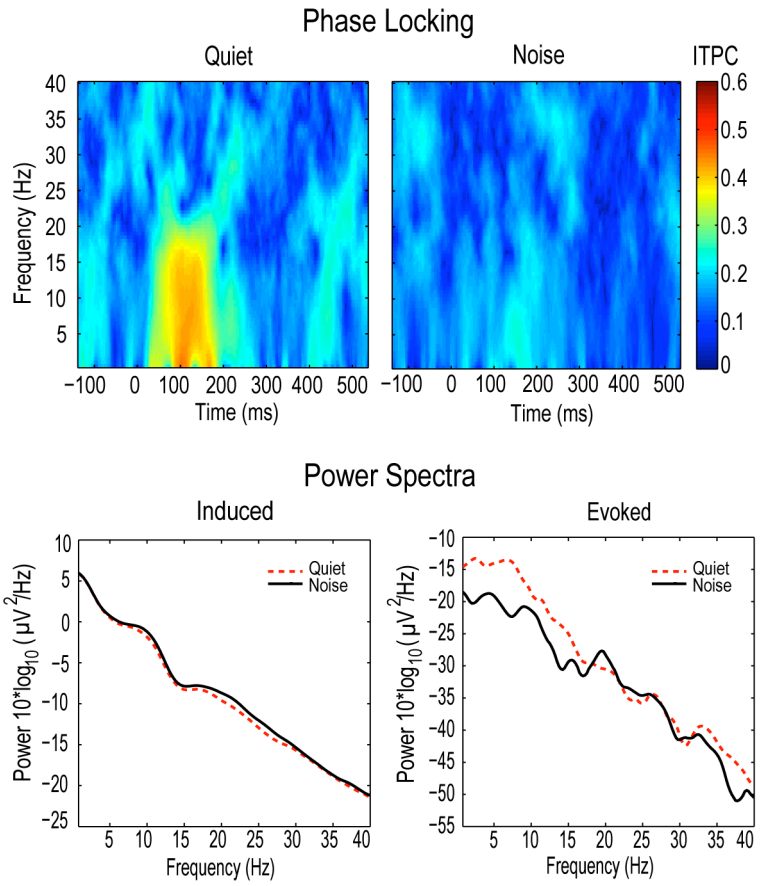

Figure 3. Top row: phase locking spectrograms for the CV syllable /bu/ in quiet and in noise at electrode Cz using inter-trial phase coherence (ITPC) that range from 0-1 and represent the strength or consistency of neural oscillations across trials averaged across participants. Bottom row: power spectra averaged from *induced* trial-by-trial responses averaged across participants (left) and *evoked* subject-by-subject ERP responses (right) for the CV syllable /bu/ in quiet (red dashed) and in noise (blue) at electrode Cz.

C. Correlational Results

Table 3 shows the correlation coefficients for relationships between inter-trial neural phase locking in the delta, theta, and alpha frequency bands and the N1-P2 latency and amplitude measures. Scatterplots depicting the relationship between neural phase locking in the alpha band and N1-P2 latencies and amplitudes are shown in Figure 4. Significant correlations were found between phase locking values and N1 latency in delta ($r = -0.586$, $p < 0.01$), theta ($r = -0.521$, $p < 0.05$), and alpha ($r = -0.510$, $p < 0.05$)

frequency bands as well as N1 amplitude in delta ($r = -0.780$, $p < 0.001$), theta ($r = -0.765$, $p < 0.001$), and alpha ($r = -0.720$, $p < 0.001$) frequency bands. Similarly, the correlation analysis revealed significant associations between phase locking values and P2 latency in delta ($r = -0.468$, $p < 0.05$), theta ($r = -0.575$, $p < 0.01$), and alpha ($r = -0.586$, $p < 0.01$) frequency bands as well as P2 amplitude in delta ($r = 0.666$, $p < 0.01$), theta ($r = 0.612$, $p < 0.01$), and alpha ($r = 0.599$, $p < 0.01$) frequency bands. These results clearly demonstrated that as inter-trial neural phase locking decreased, N1-P2 latencies tended to increase and their amplitudes tended to decrease. Thus, the variations in inter-trial phase locking were able to account for the modulatory trends in the evoked N1 and P2 responses in the two listening conditions.

Table 3. Correlation coefficients for relationship between phase locking values and N1 and P2 latency and amplitude values in response to the CV syllable /bu/ at electrode Cz.

	N1		P2	
Frequency Band	Latency	Amplitude	Latency	Amplitude
Delta	-0.586**	-0.780***	-0.468*	0.666**
Theta	-0.521*	-0.765***	-0.575**	0.612**
Alpha	-0.510*	-0.720***	-0.586**	0.599**

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

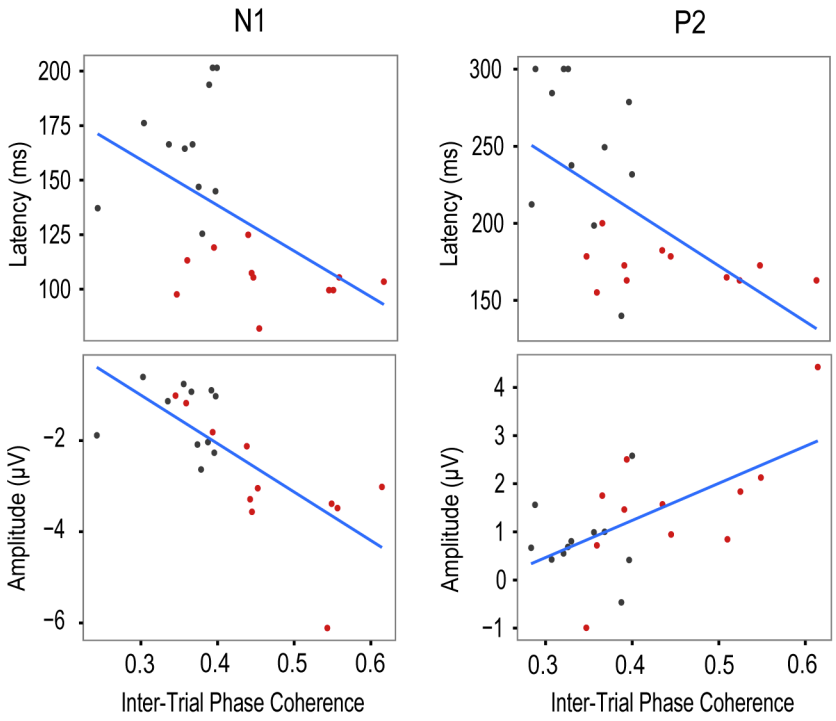

Figure 4. Scatterplots with line of best fit depicting correlations between neural inter-locking phase coherence and N1 and P2 latency and amplitude in the quiet (red) and noise (black) listening conditions in the alpha frequency band for each participant at electrode Cz. Note that it is expected that the reversed polarities between the N1 and P2 amplitudes invert the phase locking - amplitude correlations between those two measures.

IV. Discussion

This study aimed to investigate whether variations in the N1-P2 responses in quiet and in a speech-babble background noise are associated with reduced inter-trial neural phase locking to speech stimuli. Results showed that the N1-P2 latency and amplitude measures as well as inter-trial neural phase locking were affected by background noise compared to the quiet listening condition. Moreover, there were significant correlations between these measures, which suggests that neural phase locking as assessed in ITPC represents an additional tool for examining AERP latency and amplitude changes caused by the presence of background noise.

Our results on the effects of background noise on the amplitude and latency of N1 and P2 responses support previous findings (Billings et al., 2011; Martin et al., 1997; Martin & Stapells, 2005; Martin et al., 1999; Whiting et al., 1998). On average, when the four-talker speech babble at -3 dB SNR was introduced as background noise, N1 and P2 latencies increased and N1 amplitude decreased in response to the CV syllable /bu/.

Our data further illustrate that evaluating the trial-by-trial coherence in terms of phase locking allows us to gain insights into the underlying neural generators of the AERP. The results confirm that synchronized neural oscillations in delta (0.5-4 Hz), theta (4-8 Hz), and alpha (8-13 Hz) frequency bands are important contributors to the noised-induced changes in the N1-P2 complex response to speech stimuli (Başar et al., 1999; Başar, Başar-Eroglu, Karakaş, & Schürmann, 1999; Başar et al., 2000; Demiralp et al., 1999; Luck, 2014; Pantev et al., 1994; Zhang et al., 2011). However, correlations between ITPC measures and N1-P2 amplitudes appear to be stronger than those between ITPC measures and N1-P2 latencies, which could be due to differences in the metric scales for the quantities involved (i.e., microvolts vs. ms). Alternatively, the weaker correlation between latency and ITPC may point to another mechanism contributing to the variations in timing of the AERP responses across the two listening conditions.

The noise-induced changes in cortical neural synchrony have important implications for the interpretation of the averaged N1-P2 waveforms. The fact that the peak phase locking value for the N1-P2 responses across the two conditions was smaller than 0.5 in the three frequency bands suggests that the time-locked N1-P2 responses may show substantial amounts of jittering across trials even in the quiet listening condition.

The presence of babble speech noise would produce both energetic and informational masking (Kid, Mason, Richards, Gallun, & Durlach, 2007), which can reduce the detectability of the target syllable in each single trial and affect consistency of onset detection of the syllable from trial to trial. As ERP waveforms are averaged in the time domain under the time-locking and phase-locking assumption for evoked responses, the inconsistent phases of the individual cortical rhythms at the time points of peak ERP components could mask important information related to the encoding of speech at the cortical level (Luck, 2014). Our data demonstrate that reduced ITPC in the delta, theta, and alpha oscillatory bands may be partially responsible for AERP latency increases and amplitude decreases in averaged AERP waveforms recorded in response to speech in noise. In other words, the presence of background noise implies non-optimal information transmission between cortical networks related to auditory and speech processing. In this regard, trial-by-trial time-frequency analysis can provide novel and insightful information regarding the effects of noise on N1-P2 responses to auditory stimuli.

A cautionary note is necessary here that correlation does not imply causation. Our data seem to support the interpretation that the lack of phase locking across trials is one contributing factor to the noise-induced N1-P2 latency delay and amplitude reduction. But we need to point out that phase synchronization as measured in ITPC in itself is not proof that the AERP is generated by phase-resetting of the EEG. Both AERP and ITPC measures reflect phase-locked activity. The "additive" and "phase resetting" models are not necessarily exclusive of each other (See Bastiaansen et al., 2012 for a discussion of the "shared generator" hypothesis). The fact that the induced power spectrum for the

quiet and noise conditions was nearly identical is clearly consistent with the phase resetting model. However, we cannot rule out an alternative explanation that would be in line with the "additive" model. That is, the target /bu/ syllable coding part of the EEG signal can be reduced by noise in each trial and the power reduction may simultaneously be offset by the neural coding of the babble speech sentences. It is well established that speech processing at segmental and suprasegmental levels and processing of amplitude- and frequency-modulated auditory signals in general would lead to oscillatory EEG power increase in multiple frequency bands (Giraud & Poeppel, 2012).

V. Conclusion

The present study aimed to address whether changes in inter-trial phase coherence were correlated with variations in cortical N1-P2 responses to speech in noise. Consistent with our prediction, the introduction of continuous speech babble background noise disrupted cortical neural synchrony in the delta, theta, and alpha frequency bands at the time points that corresponded to the N1 and P2 responses without affecting the phase locking patterns in the baseline and the averaged trial-by-trial power spectrum. Moreover, the reductions in ITPC were able to account for variations in N1 and P2 latencies and amplitudes. The fact that much weaker correlations were found between latency (as opposed to amplitude) and ITPC suggests a possibility that the noise-induced changes in N1-P2 latencies could be additionally driven by another mechanism. The results have important implications regarding the interpretation of noise-induced changes in the early encoding of speech. Future research could further investigate whether the ITPC analysis may be a feasible research and clinical tool for assessing auditory masking

effects and speech-in-noise perception in various populations with or without hearing impairment.

Chapter 3: The MMN as a Neural Correlate of Speech Perception in Noise

Sections I-V are reprinted from:
Koerner, T.K., Zhang, Y., Nelson, P., Wang, B., & Zou, H. (2016). Neural indices of phonemic discrimination and sentence level speech intelligibility in quiet and noise: A mismatch negativity study. *Hearing Research,* 339, 40-49.

I. Introduction

Speech communication often takes place in the presence of background noise, which can be difficult for hard of hearing listeners as well as many listeners with normal hearing. In recent years, there has been a surge of interest investigating noise-induced modulatory effects on cortical/subcortical responses to examine the neural networks and brain mechanisms supporting higher-level cognitive and linguistic skills (Billings, Mcmillan, Penman, & Gille, 2013; Du, Buchsbaum, Grady, & Alain, 2014; Koerner & Zhang, 2015; Mesgarani, David, Fritz, & Shamma, 2014; Wong, Uppunda, Ajith, Parrish, & Dhar, 2008). Cortical auditory event-related potentials (AERPs) are one representative method of measuring the neural coding of speech sounds in various listening conditions. In particular, the auditory mismatch negativity (MMN) response provides an objective electrophysiological measure of the neural timing and strength of pre-attentive auditory discrimination. It peaks at approximately 100-250 ms post-stimulus onset, which is typically generated when a participant's sensory memory trace of a "standard" stimulus detects a change by a less frequently occurring "deviant" stimulus in the absence of attention or any overt behavioral response (Näätänen et al., 2007). The present study attempts to address whether the MMN response is a good predictor of speech perception performance at both segmental and sentence levels in quiet and noise.

The pre-attentive cortical MMN response has been linked with behavioral speech perception in a number of studies. Representative topics include language learning and development in children (Kraus et al., 1996; Kraus & Cheour, 2000), native (Aaltonen, Niemi, Nyrke, & Tuhkanen, 1987; Christmann, Berti, Steinbrink, & Lachmann, 2014) and non-native speech perception in adults (Bidelman & Dexter, 2015; Brunellière, Dufour, & Nguyen, 2011; Näätänen et al., 1997; Winkler et al., 1999; Zhang et al., 2009), the effects of hearing loss (Kraus et al., 1995; Oates, Kurtzberg, & Stapells, 2002) and cochlear implantation (Groenen, Snik, & van den Broek, 1996; Kraus et al., 1993; Oates et al., 2002), and neural plasticity in auditory training (Kraus et al., 1995; Tremblay, Kraus, Carrell, & McGee, 1997; Tremblay, Kraus, & McGee, 1998; Zhang et al., 2009). Studies have shown that the MMN responses for phonetic discrimination in quiet can predict first- and second- language attainment in children (Garcia-Sierra et al., 2011; Jansson-Verkasalo et al., 2004; Jackoby et al., 2011; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005). There is also evidence that pre-attentive speech perception in noise results in MMN amplitude decreases and latency increases when compared to quiet conditions (Kozou et al., 2005; Martin & Stapells, 2005; Martin et al., 1999; Muller-Gass, Marcoux, Logan, & Campbell, 2001). These noise-induced changes in the MMN response for detecting phonemic changes are associated with decrements in behavioral measures of discriminatory accuracy and increases in reaction time (Martin & Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001). However, to our knowledge, no adult studies have examined brain-behavior relationships between changes in the pre-attentive MMN at the

segmental level and performance on sentence-level word recognition performance across quiet and noise conditions.

In addition to the conventional ERP latency and amplitude measures, a recent trend in neurophysiological studies is the development of sophisticated time-frequency analyses to examine the role of various neural oscillation frequency bands of the EEG signal in the generation of AERP waveforms. These cortical oscillations are thought to modulate neural excitability and timing, which enables information exchange between cortical processes that are responsible for sensory and cognitive events (Başar et al., 1999; Klimesch et al., 2007; Koerner & Zhang, 2015; Luck, 2014; Makeig, Debener, Onton, & Delorme, 2004; Sauseng et al., 2007; Zhang et al., 2011). In particular, several studies have revealed the contribution of the theta frequency band (4-8 Hz) in driving the neuronal generation of the MMN in frontal and temporal areas (Choi et al., 2013; Fuentemilla et al., 2007; Hsiao, Wu, Ho, & Lin, 2009; Ko et al., 2012). Collectively, these studies show that neural generation of the MMN response are accompanied by phase alignment and power modulation of theta band activity. In the literature, the theta activity is proposed to be associated with several other cognitive functions including memory encoding, retrieval, and maintenance (Klimesch, Freunberger, Sauseng, & Gruber, 2008; Ward, 2003). Although previous studies have revealed prolonged latency and reduced amplitude in the MMN response due to the presence of background noise (Kozou et al., 2005; Martin & Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001), it remains unknown how noise may modulate MMN spectral power in the theta band.

The present study aimed to investigate the effects of speech-babble background noise on the pre-attentive cortical processing of consonant and vowel changes by analyzing MMN latency, amplitude, and EEG spectral power measures. It is well established that the MMN responses can show high inter- and intra-subject variability in amplitude and latency (Kraus et al., 1995; Kurtzberg, Vaughan, Kreuzer, & Fliegler, 1995; Lang et al., 1995; Martin, Tremblay, & Korczak, 2008; Martin & Stapells, 2005; Martin et al., 1999; Näätänen, 1995; Näätänen et al., 2007; Stapells, 2002). This study was designed to take individual variability into account and investigate whether the objective neurophysiological measures in response to consonant and vowel phonemic contrasts in a double-oddball paradigm (Xi et al., 2010) can predict sentence-level speech intelligibility performance across quiet and noise listening conditions. The double oddball paradigm is a modified version of the conventional MMN protocol, in which the presentation trials for the single deviant stimulus are shared by two deviants (e.g., a consonant contrast and a vowel contrast) at equal probability of occurrence. Thus this paradigm allows the investigation of two MMN responses, one for each deviant stimulus, during the same recording session. Animal and human studies examining the neural processing of speech in noise have revealed differential effects of noise on consonant and vowel stimuli, such that the neural responses to steady-state vowel stimuli are more robust in noise than those to more transient, aperiodic consonant stimuli (Cunningham, Nicol, King, Zecker, & Kraus, 2002; Russo, Nicol, Musacchia, & Kraus, 2004; Shetake et al., 2011; Song, Skoe, Banai, & Kraus, 2011). Additionally, behavioral and neurophysiological research suggests that consonant and vowel stimuli may be processed

by separate neural mechanisms in the auditory cortex (Caramazza, Papagno, & Ruml, 2000; Carreiras, Vergara, & Perea, 2009; Fogerty & Humes, 2012; Fogerty, Kewley-Port, & Humes, 2012; Kewley-Port, Burkle, & Lee, 2007; Liberman & Mattingly, 1985; Miceli, Capasso, Benvegnù, & Caramazza, 2004). Thus the use of a double-oddball paradigm would allow us to test the differences in neural sensitivity to vowel and consonant changes across the quiet and noise conditions and their relative contributions to higher-level behavioral performance in sentence recognition.

We hypothesized that the introduction of background noise would result in increases in MMN latency and decreases in MMN amplitude, which would be accompanied by reduced spectral power in the theta band. In addition, noise would differentially affect the cortical processing of the consonant and vowel changes, such that the pre-attentive detection of the consonant change would be more vulnerable to disruption in noise than the vowel change. We further hypothesized that at least some of the MMN measures would be able to predict higher-level behavioral sentence recognition.

II. Methods

A. Subjects

The participants in the study were 15 individuals (mean age = 22.6 years, age range = 19 – 32 years, 5 males, 10 females) with normal hearing (as shown in standard audiological assessment with hearing thresholds < 25 dB HL for pure tones from 0.25 to 8 kHz) and no history for speech, language, or cognitive difficulties. All participants were right handed and were native speakers of American English. The Human Research

Protection Program at the University of Minnesota approved the research protocol and all participants provided informed consent prior to beginning the study.

B. Stimuli

Stimuli for ERP measures

The consonant-vowel (CV) syllables, /ba/, /da/, and /bu/, were synthesized with the HLsyn software program (Sensimetrics Corporation, USA) using a 10 kHz sampling rate (Koerner & Zhang, 2015a). All the syllables were 170 ms in duration with a steady fundamental frequency of 100 Hz and a steady F4 at 3300 Hz. The HLsyn software generated formant transitions in the first 50 ms of the CV syllables with onset frequencies at 328 Hz, 1071 Hz, and 2298 Hz respectively for F1, F2, and F3 of the /ba/ sound. For /da/, the F1, F2, and F3 onset frequencies were 362 Hz, 1832 Hz, and 2540 Hz, and for /bu/, the formant onset frequencies were at 230 Hz, 900 Hz, and 2480 Hz. The steady center F1, F2, and F3 frequencies for the vowel portion (50-170 ms) of the /ba/ and /da/ syllables were 674 Hz, 1140 Hz, and 2350 Hz. The steady center F1, F2, and F3 frequencies for the vowel portion of /bu/ were 320 Hz, 860 Hz, and 2620 Hz, respectively. The background noise used in this study was a four-talker speech babble noise that was adopted from the Quick Speech In Noise Test (Quick-SIN) (Niquette, Gundmundsen, & Killion, 2001). All of the CV syllables and the noise stimuli were resampled at 44.1 kHz and were normalized to create a -3 dB SNR using Sony SoundForge 9.0 (Sony Creative Software, USA) (Koerner & Zhang, 2015).

Stimuli for Behavioral Measures

The CV syllables were utilized in an active listening condition (described below) in order to obtain percent correct phoneme change-discrimination as well as response reaction time. Participants also listened to IEEE sentences (IEEE, 1969) in both the quiet and four-talker speech babble masker conditions to obtain sentence recognition scores.

C. Procedure

All testing was conducted in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems). The reported EEG data were taken from a larger scale study with a two-hour recording session including both passive and active listening conditions (Koerner, Zhang, & Nelson, 2013). The CV syllable stimuli were presented using EEvoke software (ANT Inc., Netherlands) via bilateral Etymotic ER-2 insert headphones. The speech signal was presented at 60 dB SL relative to the individual participants' hearing thresholds at 1 kHz (Koerner & Zhang, 2015). Participants were presented with two different conditions: signals in quiet and signals in a four-talker speech babble noise at a -3 dB SNR in both EEG and behavioral tests.

Stimulus presentation order in the double-oddball paradigm was pseudo-randomized so that no two deviants were presented in succession and that no blocks began with a deviant stimulus. Two speech contrasts were included in the double-oddball paradigm: a vowel change (from /ba/ to /bu/) and a consonant change (from /ba/ to /da/). The vowels /a/ and /u/ are both back vowels, which differ primarily in F1 in the steady vowel portion (Ladefoged, 2006). On the other hand, the /b/ and /d/ consonants, represent transient differences in place of articulation cued by second and third formant frequency transitions, which are in a frequency range commonly affected by hearing

impairment (Ladefoged, 2006; Miller & Nicely, 1955).The standard stimulus /ba/ had a probability of occurrence of 0.75 and the two deviant stimuli, /da/ and /bu/, each had a probability of occurrence of 0.125. Each condition consisted of 10 blocks for a total of 832 trials for standard stimuli and 104 trials for each deviant stimulus. Both /da/ and /bu/ were presented as standard stimuli in 4 separate alternating blocks of 30 repetitions for a total of 120 repetitions of each stimulus. Neural responses to these "standard" presentations of /da/ and /bu/ were subtracted from the same syllables when they were presented as deviant stimuli in the double-oddball paradigm. Kraus et al. (1995) advocated the use of a "deviant alone" recording session when there might be large acoustic confounds (in our case, consonant vs. vowel contrasts) in the interpreting the MMN data. This method of ERP subtraction is used to obtain the "identity MMN," which controls for potential acoustic confounds from simple deviant-minus-standard subtraction (Zhang et al., 2005; Pulvermüller & Shtyrov, 2006). The interstimulus interval (ISI) was randomized from trial to trial within the range of 600 – 700 ms. The inter-block interval was 5 s. The quiet and noise conditions were counter-balanced across subjects to reduce potential test order effects. During the MMN recording session, participants were instructed to relax, minimize excessive movements (including blinking), and stay awake while watching a muted movie of their choice with subtitles and ignoring the auditory stimuli played through the headphones.

Behavioral responses for syllable detection were recorded from a separate discrimination task using the same trials and presentation levels as in the double-oddball paradigm. Unlike the MMN recording session, no muted movie was presented. Instead,

the participants were asked to focus their attention on detecting sound changes and make a button-press response using a desktop keyboard whenever they heard a deviant stimulus. This was followed by an additional 30-minute behavioral session where speech intelligibility responses to randomized lists of IEEE sentences presented through TDH-39 headphones were recorded. During this speech intelligibility test session, participants were instructed to repeat out loud word by word, as best as they could, the IEEE sentences they heard. Their word-by-word responses were then evaluated by a certified audiologist for recognition accuracy.

D. Data Analysis

ERP measures

Continuous EEG data were recorded using the Advanced Neuro Technology EEG System (Advanced Source Analysis version 4.7) and a 64-channel Ag AgCl electrode WaveGuard cap with a REFA-72 amplifier (TMS International BV) (bandwidth = 0.016-200 Hz, sampling rate = 512 Hz). The average impedance of electrodes was below 5 kOhms. ERP waveform analysis was completed offline in BESA (Version 6.0, MEGIS Software GmbH, Germany). The EEG data were bandpassed at 0.5-40 Hz. The ERP epoch length consisted of a 100 ms prestimulus baseline and a 700 ms poststimulus interval. Automatic artifact rejection criteria were set at ± 50 μV. The MMN was analyzed with a common average reference at the Cz electrode. Based on the grand average waveforms in the quiet and noise conditions, MMN peak latency, relative to the pre-stimulus interval, was assessed within the time window of 100 -300 ms. The MMN

amplitude quantification used an integration (averaging) window of 10 ms centered at peak (Rao, Zhang, & Miller, 2010).

Spectral power analysis was completed using the subtracted MMN waveform at electrode Cz with the *newtimef* function in EEGLAB (Delorme & Makeig, 2004). A short-term Fourier Transform (STFT) with Hanning window tapering (Koerner & Zhang, 2015), which is recommended for the analysis of low frequency activities, was adopted to extract the logarithmic spectra power for the theta band. The modified STFT method in EEGLAB used overlapping sliding windows that are adapted to the target frequency bins to overcome limitations due to the use of fixed windows in conventional analysis. Zero-padding was applied for short epochs with insufficient sample points for the Fourier transform. The time window used for our time-frequency analysis represented the entire analysis epoch, including the pre-stimulus baseline from -100 to 700 ms, and estimated frequencies were from 0.5-40 Hz with a step interval of 0.5 Hz (Koerner & Zhang, 2015).

All statistical analyses from both the AERP and behavioral portions of the study were analyzed in R (R Core Team, 2014). A repeated measures analysis of variance (RM-ANOVA), with $\alpha = 0.05$, was conducted to examine the statistical significance of stimulus type (/da/ or /bu/), condition (quiet or noise), and any potential interactions (stimulus x condition) on MMN latency, amplitude, and spectral power recorded at electrode Cz. Where significant interaction effects were observed, tests of simple main effects and post hoc two-tailed t-tests for selected factors of interest were also conducted to evaluate how consonant and vowel changes were processed differently in quiet and in noise.

Behavioral Measures

In the phoneme discrimination test, the percent correct scores and reaction time for the detection of consonant and vowel changes in the double-oddball paradigm were obtained from the button-press responses recorded during the quiet and noise conditions. A RM-ANOVA was completed to investigate the statistical significance of stimulus type (/da/ or /bu/), condition (quiet or noise), and any potential interactions (stimulus x condition) on behavioral discrimination accuracy and reaction time.

In the speech intelligibility test, the IEEE lists consisted of 10 low context sentences with 5 key words in each sentence. For both the quiet and noise conditions, two sentence lists of 50 key words were presented from one female and one male talker for a total of 100 key words per listening condition. Participants repeated key words from 4 randomized lists of 10 behavioral IEEE sentences (IEEE, 1969) and percent correct scores for key words were recorded. An additional RM-ANOVA was carried out to examine the significance of listening condition (quiet vs. noise) on IEEE sentence recognition.

Brain-Behavior Correlations

Using R (R Core Team, 2014) and the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2016), linear mixed-effects models were developed to examine whether the MMN measures were predictive of behavioral speech perception at both the syllable- and sentence- level. LME models are extension of linear regression models, which are particularly useful in settings involving repeated measures with coefficients that can vary with respect to one or more grouping variables. Regression analyses have

previously been used to examine whether combinations of ERP data collected from individual subjects are viable predictors of behavioral speech perception across different listening conditions (Billings et al., 2013; Billings, Penman, Mcmillan, & Ellis, 2015). Data transformations included re-scaling MMN latency and behavioral reaction time values. Additionally, the behavioral percent correct phoneme detection and sentence recognition variables were log transformed to account for skewness in the data that was observed during visual analysis.

Participants were used as a "by-subject" random effect in each linear mixed-effect model. Listening condition (quiet vs. noise) and stimulus (/bu/ vs. /da/) were included as blocking variables. We then added MMN latency, amplitude, and theta power as fixed effects in order to predict percent correct phoneme detection and reaction time. Similar models were developed to evaluate whether the MMN was able to predict sentence-level perception. In implementing the LME models, we first transformed the original data to alleviate large skewness and difference in magnitudes. We adopted the "by-subject" random effects, and two block variables: condition (quiet vs. noise) and stimulus (/bu/ vs. /da/). The significance of each fixed effect in predicting each behavioral outcome measure was assessed with $\alpha = 0.05$.

III. Results

Analysis of behavioral data revealed noise-induced increases in reaction time for phoneme discrimination as well as reductions in percent correct phoneme discrimination and sentence recognition (Table 4). In the ERP analysis, the presence of speech-babble background noise increased MMN latency and decreased MMN amplitude (Fig. 5 and

Table 5), which was accompanied by decreased MMN spectral power in the theta frequency band (Table 5). Linear mixed-effects regression analysis revealed several significant neural predictors of behavioral perception at the phoneme and sentence-level (Table 6).

A. Behavioral Phoneme Discrimination and Sentence Recognition

RM-ANOVA results revealed that the presence of background noise significantly prolonged behavioral reaction time during phoneme discrimination ($F(1,14) = 41.4$, $p < 0.001$) (Table 4). Additionally, reaction time was significantly shorter during behavioral discrimination of the vowel change (from /ba/ to /bu/) compared to the consonant change (from /ba/ to /da/) ($F(1,14) = 24.95$, $p < 0.001$). Performance accuracy in noise was significantly poorer compared to the quiet listening condition ($F(1,14) = 179.1$, $p < 0.001$). As expected, there was also a significant stimulus effect with better accuracy for detecting the vowel change ($F(1,14) = 232.3$, $p < 0.001$). A significant interaction between listening condition and stimulus indicated a differential effect of noise on the two deviant stimuli ($F(1,14) = 223.2$, $p < 0.001$) with a larger noise-induced decline in behavioral performance for detecting the consonant change. Post hoc comparisons of the two speech syllables revealed that the introduction of background noise significantly decreased percent correct detection of /da/ ($t(14) = 15.50$, $p < 0.001$) and had a smaller, but still significant impact on the detection of /bu/ ($t(14) = 3.37$, $p < 0.05$). Noise-induced reduction in performance was also seen in behavioral sentence recognition ($F(1,14) = 72.15$, $p < 0.001$).

Table 4. Means (standard error) for behavioral percent correct detection of vowel (/bu/) and consonant (/da/) changes, behavioral reaction time for vowel (/bu/) and consonant (/da/) changes, as well as percent correct behavioral sentence recognition performance.

Behavioral Measure	Listening Condition	
	Quiet	Noise
/bu/ detection (%)	98.20 (0.6)	91.79 (2.22)
/da/ detection (%)	92.68 (1.63)	34.05 (4.03)
Sentence recognition (%)	99.27 (0.37)	67.53 (3.77)
/bu/ response time (ms)	468.04 (10.01)	551.51 (16.83)
/da/ response time (ms)	510.98 (12.61)	621.91 (12.10)

B. Amplitude and Latency Measures for Averaged MMN Response

RM-ANOVA results showed significant effects of speech babble background noise on MMN latency ($F(1,14) = 29.43$, $p < 0.001$) and amplitude ($F(1,14) = 32.52$, $p < 0.001$) (Table 5 & Figure 5). The presence of noise led to a significant increase in the MMN latency and decrease in amplitude during passive speech perception. Additionally, there were significant differences in latency ($F(1,14) = 17.84$, $p < 0.001$) between the two deviant syllables /da/ and /bu/ across the quiet and noise conditions. The MMN for /da/ peaked later than that for /bu/. A significant interaction effect was found between stimulus and condition for the MMN amplitude measure ($F(1,14) = 18.77$, $p < 0.001$), indicating that noise had a differential effect on the neural processing of the consonant and vowel change. Further t-tests showed that a large reduction in MMN amplitude occurred for the AERP recorded in response to the more salient CV syllable /bu/ ($t(14) = -6.30$, $p < 0.001$) but not in response to /da/ ($t(14) = -0.65$, $p = 0.528$). Conversely, significant increases in MMN latency occurred for both the CV syllable /bu/ ($t(14) = -3.18$, $p < 0.01$) and /da/ ($t(14) = -4.97$, $p < 0.001$). In addition, MMN responses to /da/

and /bu/ in quiet were significantly different in terms of amplitude ($t(14) = 3.61$, $p < 0.01$) and latency ($t(14) = 3.92$, $p < 0.01$).

Table 5. MMN mean latency (ms), amplitude (μV), and power in the theta band (dB) (standard error) in response to the CV syllable vowel change (/bu/) and consonant change (/da/) at electrode Cz in quiet and in noise.

	Latency (ms)		Amplitude (μV)		Spectral Power (dB)	
	Quiet	Noise	Quiet	Noise	Quiet	Noise
/bu/	171.88 (10.53)	236.98 (14.80)	-2.11 (0.30)	-0.40 (0.12)	-10.77 (1.30)	-19.53 (1.46)
/da/	228.26 (13.79)	293.10 (11.45)	-1.16 (0.20)	-1.02 (0.18)	-17.79 (1.28)	-18.57 (1.55)

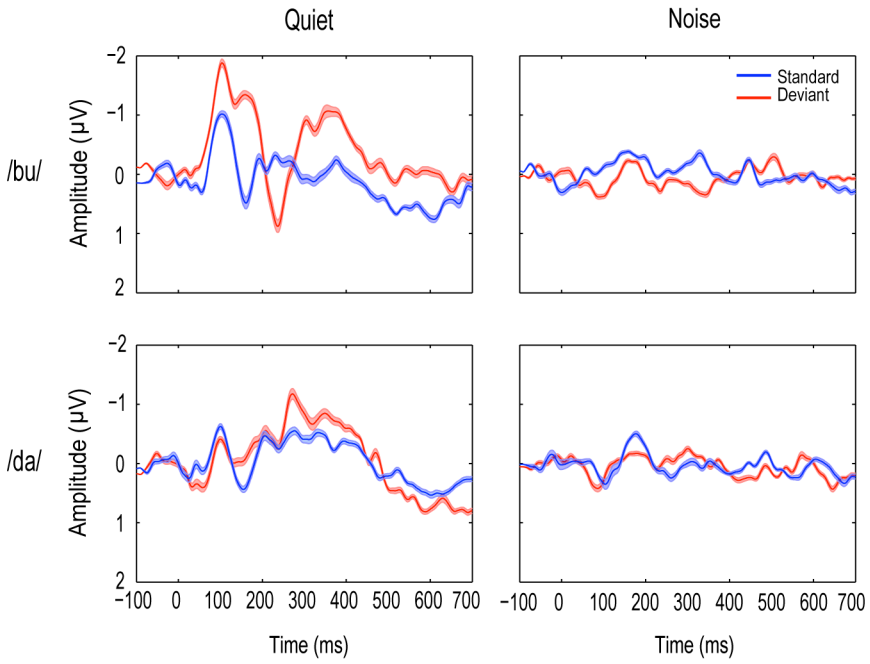

Fig. 5. Grand mean AERP waveforms averaged across participants in response to the standard (blue) and deviant (red) stimuli (top row: /bu/, bottom row: /da/) for quiet and noise conditions at electrode Cz with shaded standard error envelopes.

C. MMN Spectral Power

As predicted, MMN spectral power in the theta band was significantly reduced in background noise compared to the quiet condition across the two deviant CV stimuli ($F(1,14) = 19.37$, $p < 0.001$) (Table 5). There was also a main effect of stimulus with /da/

showing significantly lower theta power than /bu/ ($F(1,14) = 11.21$, $p < 0.01$). A significant stimulus by condition interaction was found ($F(1,14) = 10$, $p < 0.01$). Further t-tests showed a large reduction in response to the CV syllable /bu/ in noise ($t(14) = 5.60$, $p < 0.001$) but not in response to /da/ ($t(14) = 0.45$, $p = 0.661$). In addition, theta power in response to /da/ was significantly different from that of /bu/ in quiet ($t(14) = -4.33$, $p < 0.001$).

D. Linear Mixed-Effect Model Results

Linear mixed-effects regression analysis revealed that MMN latency ($F(1,40) = 7.86$, $p < 0.01$) and theta power ($F(1,40) = 6.61$, $p < 0.05$) were significant predictors of behavioral phoneme detection accuracy across conditions and stimuli. Additionally, MMN amplitude for phoneme detection showed a trend of approaching significance as a predictor of behavioral accuracy ($F(1,40) = 3.10$, $p = 0.086$). Linear mixed-effects regression analysis also revealed that theta power was significantly correlated with MMN latency ($F(1,42) = 5.61$, $p < 0.05$) and amplitude ($F(1,42) = 11.28$, $p < 0.01$). In contrast, there was no significant correlation between any of the MMN measures and the behavioral reaction time data (MMN latency ($F(1,40) = 0.00$, $p = 0.996$), amplitude ($F(1,40) = 0.00$, $p = 0.961$), and theta power ($F(1,40) = 0.37$, $p = 0.547$)). In the analysis for behavioral sentence-level scores, the MMN amplitude in response to /bu/ was the only significant predictor of behavioral performance ($F(1,11) = 7.21$, $p < 0.05$) (see Table 6 for a summary of regression model outputs for each behavioral outcome measure).

As the residual plots of all models showed no sign of any significant trend or heteroscedastic variance, we do not expect any potential improvement from the use of

non-linear models. As all the residuals appear normally distributed, we also do not expect better results (in terms of finding more or stronger brain-behavior correlations in the current data set) from generalized linear models.

Table 6. F-statistics for fixed effects from linear mixed-effects regression models for each behavioral measure

Variable	Percent Correct Phoneme Detection	Phoneme Detection Reaction Time	Percent Correct Sentence Recognition (/bu/)	Percent Correct Sentence Recognition (/da/)
Intercept	161.51***	4199.98***	431.41***	335.12***
Condition	131.68***	61.92***	291.32***	247.69***
Stimulus	114.20***	21.05***	-	-
Latency	7.86**	0.000	1.24	0.44
Amplitude	3.10	0.002	7.21*	0.41
Theta Power	6.61*	0.368	0.46	1.50

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

IV. Discussion

This study was designed to examine the effects of background noise on the MMN recorded in response to a consonant versus vowel change and to determine whether noise-induced variations in these objective cortical measures are able to predict segmental- and sentence-level behavioral speech recognition in noise.

A. Noise-induced decreases in behavioral accuracy and increases in response time

Our overall results are consistent with previous studies that have examined the effects of noise on the behavioral perception of consonant and vowel stimuli (Miller & Nicely, 1955; Parikh & Loizou, 2005; Phatak & Allen, 2007; Pickett, 1957). Although all behavioral measures were negatively impacted by the introduction of background noise, our data showed consistently lower performance for detecting a consonant change across the two listening conditions. It is well known that the important perceptual cues

contained in consonants differ from those in vowels (Ladefoged, 2006). Consonants tend to be aperiodic, weaker in acoustic energy, and have faster formant transitions that may induce more susceptibility to the deleterious effects of noise compared to more salient and sustained vowel cues. Parikh and Loizou, (2005) reported that even at poor SNRs (signal-to-noise ratios), vowels still had relatively intact F1 cues and partial F2 cues available for vowel identification, and although performance was still high in noise, several features known to be associated with consonant place of articulation identification were significantly impacted.

This difference in acoustic susceptibility to noise masking implies that the relative importance of consonant and vowel categories to overall speech perception may shift in the presence of noise. It has been shown that low- and high-frequency acoustic landmarks in consonant sounds provide useful information for word boundary segmentation and accurate sentence recognition, and that disruption of these landmarks by background noise produces errors in the perception of voicing and place of articulation cues (Li & Loizou, 2008). In fact, when Owren and Cardillo (2006) replaced consonant or vowel segments with silence in a same/different word-meaning task, they found that listeners were better able to perceive words with consonant-only information compared to vowel-only stimuli. However, several studies have shown that vowels tend to contribute more than consonants to sentence intelligibility (Cole, Yan, Mak, Fanty, & Bailey, 1996; Fogerty & Kewley-port, 2009; Kewley-Port et al., 2007), potentially due to contributions from amplitude envelope and temporal fine structure cues contained within vowel segments in meaningful, sentence-level contexts compared to word recognition

tasks (Fogerty & Humes, 2012). For instance, when consonant or vowel segments were removed from sentence stimuli and replaced with noise, both Cole (1996) and Kewley-Port et al. (2007) found superior sentence recognition performance when listeners had access to vowel-only information compared to consonant-only information, but showed no differences in performance on a word intelligibility task (Fogerty & Humes, 2010). Although these studies did not compare quiet and noise listening conditions, they revealed that consonants and vowels might play different roles in speech understanding under different linguistic contexts. Examining how background noise affects the processing of consonants and vowels and how these differential effects are reflected at the cortical level is important for understanding underlying causes of variability and decreased performance during speech perception in noise.

B. Differential effects of noise on the neural processing of consonants and vowels

Consistent with the behavioral results, our neurophysiological data revealed that when stimuli were presented in a double-oddball paradigm in quiet, the MMN responses to the consonant change were weaker than those to the vowel change. It is known that measures of neural processing time and magnitude of cortical activation in response to a target stimulus are dependent on the magnitude of deviation from the standard stimulus (Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1989; Pakarinen et al., 2013; Sams, Paavilainen, Alho, & Näätänen, 1985). Since our behavioral results showed significantly poorer performance for detection of the consonant change compared to the vowel change, a possible straightforward explanation for the same patterns in the behavioral data and

MMN data in quiet is simply that the /ba/-/da/ contrast represented a smaller, or more subtle, acoustic change than the competing /ba/-/bu/ contrast.

However, the acoustics-based explanation has difficulty in accounting for noise-induced MMN changes, which appeared to show a different pattern from the behavioral data. While the overall pattern of noise-induced latency increase and amplitude decrease in MMN to both deviants in our study replicated previous findings (Kozou et al., 2005; Martin & Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001), the background noise appeared to affect the detection of the vowel change more than the detection of the consonant change. Similar results were provided by Niemitalo-Haapola et al. (2015), who found a larger noise-induced MMN reduction in response to /e/-/i/ vowel changes, such that no significant MMN for a vowel change was recorded in noise, compared to a reduced, yet still present MMN to /p/-/k/ consonant changes in normally developing toddlers. Why would the background noise have a greater influence on the MMN for the more salient speech contrast? A possible explanation is that the larger noise-induced effects on pre-attentive neural processing of the more salient vowel cue could arise due to the need to internally resolve consonant information in background noise when consonant and vowel changes are juxtaposed in a double-oddball paradigm. This could imply that information contained in consonant segments might contribute more to speech perception in adverse listening environments. However, there are other important factors to consider if we extend the interpretation to sentence-level processing in noise as the relative importance of consonant and vowel cues may change depending on the level of lexical information available (Fogerty & Humes, 2010; 2012).

As we did not include nonspeech control stimuli in the current study, we could not rule out the possibility that the differences between vowel and consonant processing that we observed might reflect general auditory mechanisms of acoustic processing rather than phonemic processing. Moreover, as the current study tested only one consonant contrast and one vowel contrast, it remains to be tested whether the same phenomena would hold for other vowels/consonants. The advantage of the double oddball paradigm in the present study is that it allows us to directly compare how noise affects neural sensitivity to consonant and vowel contrasts at the pre-attentive level. Previous behavioral studies have shown the relative importance of consonant and vowel segments for speech intelligibility (Cole et al., 1996; Fogerty & Humes, 2010, 2012; Fogerty & Kewley-Port, 2009; Kewley-Port et al., 2007; Owren & Cardillo, 2006). Brain research further indicates that consonants and vowels are processed by distinct neural mechanisms (Caramazza et al., 2000; Carreiras et al., 2009). Although differential effects of noise on speech sounds have been found in neurophysiological studies (Anderson, Skoe, Chandrasekaran, Zecker, & Kraus, 2010; Cunningham et al., 2002; Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2001; Russo et al., 2004; Song et al., 2011), these studies only examined responses to the consonant and vowel portion within a single CV syllable. For instance, while examining the brainstem frequency following response (FFR) to the CV syllable /da/, Russo and colleagues (2004) found that noise caused greater disruption of the transient portion of the neural response, which reflects coding of rapidly changing stimulus features that are characteristic of consonant sounds. Additionally, previous ERP studies that examined the effect of noise on speech processing only used one phonemic

contrast in an oddball paradigm, representing either a single consonant change or a vowel change (Billings et al., 2013, 2015; Koerner & Zhang, 2015; Martin et al., 1997; Martin & Stapells, 2005; Martin et al., 1999; Whiting et al., 1998). While the use of new multi-feature recording paradigms with several juxtaposed "deviant" stimulus-changes allow for a comparison of MMN responses to deviant consonant and vowel changes in direct competition (Näätänen et al., 2004; Pakarinen et al., 2013; Pakarinen et al., 2007), these studies typically only tested stimuli in quiet. The results showed that consonant changes were more difficult to detect than vowel changes (Korczak & Stapells, 2010; Pakarinen et al., 2009), which were replicated in our quiet condition.

Overall, the MMN amplitude and latency results may suggest that consonants and vowels play different roles in speech processing, which has both theoretical and practical implications for understanding theories of speech perception and developing strategies for improving speech understanding in noise. For instance, performance may be improved by making some speech sound features more accessible through rehabilitation options such as amplification from hearing aids or by training listeners to attend to important cues contained in certain speech segments. These strategies could be tested using neural measures to examine whether reductions in noise-induced cortical effects are observed during speech perception in noise after rehabilitation or training.

C. Theta power modulation and speech perception in noise

Although it has already been established that event-related cortical oscillations in the theta frequency band are associated with linguistic processing of phonemic contrasts (Jin, Diaz, Colomer, & Sebastian-Galles, 2014), little is known about the effects of noise

on theta power modulation during pre-attentive speech discrimination. Our results are consistent with previous findings that reveal the importance of the theta frequency band in the pre-attentive neural processing of auditory deviant stimuli in quiet (Choi et al., 2013; Fuentemilla et al., 2007; Ko et al., 2012). Furthermore, we demonstrated that noise significantly decreased theta power in response to both speech stimuli.

Research examining neurocognitive linguistic development in infants has also shown that measures of EEG spectral power modulation are sensitive to stimulus features (Radicevic, Vujovic, Jelicic, & Sovilj, 2008; Santesso, Schmidt, & Trainor, 2007; Zhang et al., 2011). For instance, Zhang and colleagues (2011) found that when infants passively listened to the vowel /i/ in alternating blocks of exaggerated infant-directed or unexaggerated adult-directed speech in quiet, ERPs were enhanced and EEG spectral power in the theta band was stronger in response to the more prominent vowel. Our theta power modulation data are consistent with these findings, showing stronger spectral power for the salient vowel change in quiet. In the meanwhile, the spectral power of theta band also showed that neural processing of the more prominent vowel change had larger noise-induced effects in comparison with the consonant change.

D. The MMN as a neurophysiological marker of behavior

We chose to use a -3 dB SNR in this study to purposely evoke a wide range of performance on the speech perception tasks in order to examine brain-behavior correlations. Pilot data in our lab (Koerner, Zhang, Nelson, 2013) revealed that this noise level induced a range of performance in the detection of the easier vowel change and more difficult consonant change, as well as sentence recognition, without ceiling or floor

effects. Based on previous studies showing correlations between speech-evoked neural responses and behavioral speech perception in noise (Anderson, Chandrasekaran, Yi, & Kraus, 2010; Anderson, Skoe, Chandrasekaran, Zecker, et al., 2010; Bennett et al., 2012; Billings et al., 2013, 2015; Cunningham et al., 2001; Martin & Stapells,2005; Martin et al., 1999; Muller-Gass et al., 2001; Song et al., 2011), we presumed that this range of behavioral performance would be reflected in measures of pre-attentive cortical speech processing.

As expected, our results confirmed that the MMN could serve as a neurophysiological predictor of behavioral speech perception at both the syllable level and sentence level. Our data are consistent with previous studies that have examined relationships between the speech-evoked MMN and behavioral performance at the segmental level (Martin & Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001). Using a /ba/-/da/ syllable contrast, Martin et al. (1999) examined the effects of decreased audibility on the MMN in normal hearing listeners using different high-pass noise filters and found significant correlations between noise-induced variations in behavioral phoneme-change detection sensitivity and MMN amplitude as well as behavioral reaction time and MMN latency. Our results add to this current body of knowledge by providing information about the ability of the MMN to reflect noise-induced variability in behavioral performance when measured in response to a both consonant and a vowel change, and more importantly, how it relates to sentence-level performance.

Although MMN latency, and theta power were significant predictors of percent correct phoneme detection, none of these variables significantly predicted behavioral

reaction time for phoneme detection. Martin et al. (1999) reported a significant correlation between MMN latency and behavioral reaction time; however, they reported that the correlation was weak, especially when compared to the relationship between MMN amplitude and behavioral sensitivity. These results imply that the MMN is accurate in assessing behavioral phoneme detection accuracy, but may not provide strong predictive information about timing of the conscious decision-making process. This may be explained by the fact that the MMN is an obligatory response that reflects pre-attentive auditory discrimination; it is expected that the presence of an MMN would be associated with the ability to behaviorally discriminate between stimuli. However, participants may differ in how conservative or liberal they are in making a response during an auditory discrimination task, which would greatly impact behavioral reaction time. Therefore, the active P3 response, which is thought to reflect completion in the decision making process (Picton, 1992), may be a better indicator of behavioral reaction time than the pre-attentive MMN (Martin et al., 1997; Whiting et al., 1998).

While MMN amplitude in response to /bu/ (i.e., the vowel change) was a significant predictor of sentence-level speech intelligibility scores, MMN latency, amplitude, and theta power in response to /da/ (i.e., the consonant change) were not. The inability of the MMN in response to the consonant change to predict behavioral sentence recognition could be due to the lack of significant differences between the quiet and noise listening conditions in MMN amplitude or theta power.

Our analysis revealed that theta power was a significant predictor of both MMN latency and amplitude. Similarly, significant correlations were found between theta

power and MMN amplitude in studies examining MMN attenuation in clinical populations with schizophrenia using tonal stimulus contrasts (Hong, Moran, Du, O'Donnell, & Summerfelt, 2012; Kaser et al., 2013). These relationships suggest that our noise-induced MMN changes were mediated by the strength of synchronization of event-related oscillations during pre-attentive discrimination of speech contrasts. Our data confirm that using time-frequency analysis to obtain a measure of theta power represents an additional tool for examining speech-evoked MMN and its susceptibility to noise.

E. Novelty, Limitations and Future Directions

To our knowledge, this is the first study to examine whether noise-induced changes in the MMN reflect variability in higher-level speech perception across quiet and noise conditions. Previous studies have examined relationships between AERPs and behavioral sentence recognition in noise (Bennett et al., 2012; Bidelman & Howell, 2016; Billings et al., 2013, 2015; Parbery-Clark, Marmel, Bair, & Kraus, 2011), but none have examined pre-attentive auditory change-detection as a neurophysiological predictor of behavior both in quiet and in noise. Novel results from linear mixed-effects models showed that both MMN latency and amplitude may reflect variability in behavioral performance. Additionally, this study provided novel information regarding the ability of theta power modulation across quiet and noise conditions to predict behavioral speech perception abilities, which suggests that phase alignment and/or spectral power modulation of the theta oscillation within a neural population for the generation of the MMN response can be reflected in behavioral performance. These results imply that the theta spectral power measure represents an additional tool for predicting the effects of

noise on behavioral speech perception. The use of event-related cortical oscillations allows for a deeper examination of underlying neural processes that are implicated in noise, which possibly contribute to the wide range of variability seen in measures of speech perception in noise across listeners.

These novel results have implications for the clinical utility of the MMN as an obligatory cortical measure of more "real world" speech perception abilities, as it appears as if pre-attentive cortical measures can be used to predict sentence-level behavioral performance across stimuli and conditions. The pre-attentive MMN and measures of event-related cortical oscillations could prove useful in examining speech-processing abilities in clinical populations who are unable to provide consistent or reliable behavioral responses, such as adults with cognitive impairments or infants. For instance, pediatric hearing aid fittings may employ a measure of pre-attentive cortical auditory discrimination to examine pre- to post-fitting improvements in speech processing at the phonemic level. Additionally, pre-attentive cortical responses at the segmental level could be compared pre- and post-auditory training not only to assess the success of the program, but also to predict relative improvements in higher-level speech-in-noise perception in adults. In order to determine the reliability of the MMN as a neurophysiological correlate of behavioral perception at the individual level and further the practical field, additional stimuli, noise conditions, and participant populations, such as those with hearing loss, cochlear implants, or auditory processing disorders, should be tested to examine whether these significant effects generalize to other listening situations. Moreover, since our data showed lower variance for MMN amplitude and theta power

values to the consonant change relative to the vowel change, further assessments should evaluate the predictive power of the MMN for a consonant change that shows greater variance in noise. This would add to our theoretical knowledge by allowing for an evaluation of whether neural responses to the consonant or vowel change are better predictors of individual speech performance, which may contribute to a better understanding of the relative contributions of consonant and vowel information in speech perception across different listening contexts.

A current barrier to using the MMN for any clinical assessment of speech processing is the wide range of variability in individual responses even when behavioral performance is at a fixed level (Kraus et al., 1995; Kurtzberg et al., 1995; Lang et al., 1995; Martin et al., 2008; Martin & Stapells, 2005; Martin et al., 1999; Näätänen, 1995; Näätänen et al., 2007; Stapells, 2002). Future research may evaluate whether the spectral power analysis for targeted cortical rhythms would provide a more robust measure than MMN latency or amplitude, which can have strong practical implications for assessing event-related cortical oscillations in clinical populations with speech perception difficulties and potentially monitoring intervention outcomes.

Since our behavioral auditory change-discrimination data showed differences in the effects of noise on consonant and vowel perception compared to the pre-attentive cortical responses, future research should examine how noise impacts change detection in an active listening condition in the double-oddball paradigm by analyzing the P3 responses. This would help determine the roles that attention and listening context play in the neural processing of consonants and vowels in adverse listening conditions.

Additionally, the present study only used three CV speech stimuli, representing only one consonant and one vowel change. Due to time constraints and study design, only one noise condition was tested. It is necessary to test additional stimuli and noise levels to determine how these two speech sound classes are affected by background noise at different SNRs. It is possible that the noise level may cause shifts in differential neural processing of consonant and vowel stimuli depending on attentional demand. Additionally, non-speech acoustic control stimuli (e.g., Zhang et al., 2005) should be examined to determine whether the MMN data as reported in the current study truly reflected consonant and vowel processing or general auditory mechanisms of detecting distinct acoustic changes. Furthermore, data from hearing-impaired listeners should be evaluated, as these listeners most often have reports of difficulty listening to speech in noise even after audibility is improved via amplification from hearing aids.

V. Conclusion

This study aimed to determine whether variations in cortical speech-evoked MMN latency, amplitude, and spectral power measures in response to a consonant and vowel change could predict behavioral speech perception abilities at both syllable and sentence levels across quiet and noise conditions. Results were consistent with our predictions that the introduction of background noise would increase MMN latencies as well as decrease MMN amplitude and EEG power in the theta frequency band. As expected, the speech-babble background noise had a differential effect on the neural processing of the consonant and vowel changes. On average, MMN responses to the CV syllable /da/ had longer latencies, smaller amplitudes, and less power in the theta

frequency band across quiet and noise conditions compared to /bu/. However, it was unexpected that the addition of background noise would have a greater effect on the neural processing of the more salient vowel change when compared to the quiet condition. Finally, consistent with our expectations, the objective MMN measure in a double-oddball paradigm was a significant predictor of variations in behavioral percent correct detection of segmental-level speech stimuli as well as higher-level sentence recognition. The relevance and utility of the reported measures for potential clinical applications require further studies with different populations, stimuli, and experimental conditions.

Chapter 4: Exploring the Use of Mixed-Effects Regression Models for Examining Brain-Behavior Relationships

Sections I-V are reprinted from:
Koerner, T.K. and Zhang, Y. (2017). Application of Linear Mixed-Effects Models in Human Neuroscience Research: A Comparison with Pearson Correlation in Two Auditory Electrophysiological Studies. *Brain Sciences,* 7(26), 1-11.

I. Introduction

Cognitive neuroscience research aims to explore relationships between various neural and behavioral measures to examine the underlying peripheral/central neural mechanisms in various testing conditions and subject populations. For this purpose, the bivariate Pearson correlation analysis is commonly used to examine the strength of the linear relationship between two continuous variables of interest, which can be graphically represented by fitting a least-squares regression line in a scatter plot (McElreath, 2016; Pernet, Wilcox, Rousselet, 2013). If the variables do not represent continuous data or if the relationship between the two variables is non-linear, other types of bivariate correlation tests such as Spearman or Point-Biserial correlations can be used. However, when a study involves multivariate data, the conventional correlation method only allows for the examination of one predictor and one outcome variable at a time. Even if the Pearson correlation results are adjusted for multiple comparisons or a simple multiple regression model is applied, the statistical treatment may not take into account the complex relationships and categorical grouping terms that likely exist in the multiple within-subject predictor variables (Pernet et al., 2013).

In consideration of the violation of the assumed sample independence required of bivariate Pearson correlations and the like, researchers have long argued for the necessity to apply more sophisticated statistical techniques to handle repeated measures from the

same subjects (Baayen, Davidson, & Bates, 2008; Bagiella, Sloan, & Heitjan, 2000; Magezi, 2015). The use of mixed-effects (or multilevel) models has recently captured attention in longitudinal medical research (Andersson-Roswall, Engman, Samuelsson, & Malmgren, 2010; Ard, Raghavan, & Edland, 2015; Bilgel, Prince, Wong, Resnick, & Jedynak, 2016; Cuthbert et al., 2015; Davidson & Martin, 2013; Hasenstab et al., 2015; Maneshi, Moeller, Fahoum, Gotman, & Grova, 2012; Mistridis, Krumm, Monsch, Berres, & Taylor, 2015; Pedapati et al., 2016), behavioral and social sciences research (Agresti, Booth, Hobert, Caffo, 2000; Berger & Tan, 2004; Cheung, 2008; Luger, Suls, & Vander Weg, 2014; Parzen et al., 2011) (including speech and hearing research (Billings et al., 2013, 2015; Cahana-Amitay et al., 2016; Canault, Le Normand, Foudil, Loundon, & Hung, 2016; Cunnings, 2012; Davidson & Martin, 2013; De Kegel, Maes, Van Waelvelde, & Dhooge, 2015; Evans, Chu, Aston, & Su, 2010; Gfeller et al., 2007; Haag, Roppelt, & Heppt, 2015; Hadjipantelis, Aston, Muller, & Evans, 2015; Humes, Burk, Coughlin, Busey, & Strauser, 2007; Jouravlev & Lupker, 2015; Kasisopa, Reilly, Luksaneeyanawin, & Burnham, 2016; Linck & Cunnings, 2015; Murayama, Sakaki, Yan, & Smith, 2014; Picou, 2016; Poll et al., 2013; Quene & van den Bergh, 2008; Rong, Yunusova, Wang, & Green, 2015; Stuart & Cobb, 2015; van de Velde & Meyer, 2014)), and neurophysiological and neuroimaging research (Amsel, 2011; Bornkessel-Schlesewsky et al., 2015; Bramhall, Ong, Ko, & Parker, 2015; Hsu, Lee, & Marantz, 2011; McEvoy, Hasenstab, Senturk, Sanders, & Jeste, 2015; Payne, Lee, & Federmeier, 2015; Spinnato, Roubaud, Burle, & Torresani, 2015; Tremblay & Newman, 2015; Visscher et al., 2003; Wang, Yang, Fan, Sun, & Yue, 2009; Zenon et al., 2015). Its

increasing popularity is shown in the exponential growth over the last three decades in the number of publications in the scientific literature (Figure 6).

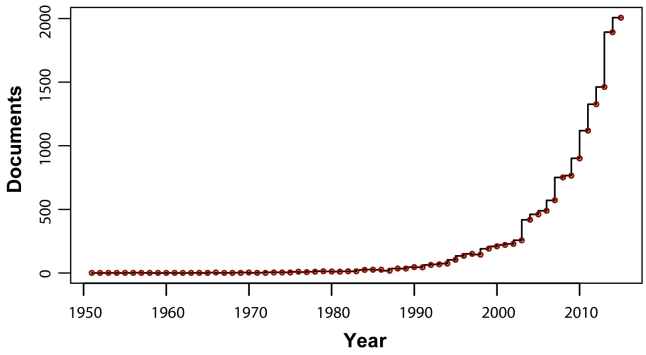

Figure 6. Number of publication documents (including original articles and reviews) from 1951 to 2016 that contain the keyword "linear mixed-effects model." Literature search was conducted with Elsevier's Scopus database (Scopus, n.d.).

Data analysis using mixed-effects regression models allows for the examination of how multiple variables predict an outcome measure of interest beyond what a simple multiple regression model can handle (McElreath, 2016; Baayen et al., 2008; Bagiella et al., 2000; Magezi, 2015). In addition to the fixed effects in a conventional multiple regression model, a mixed-effects model includes random effects associated with individual experimental units that have prior distributions. Thus mixed-effects models are able to represent the covariance structure that is inherent in the experimental design. In particular, the linear and generalized linear mixed-effects models (LME or GLME), as implemented in popular software packages such as R, prove to be a powerful tool that allows researchers to examine the effects of several predictor variables (or fixed effects) and their interactions on a particular outcome variable while taking into account grouping factors and the existing covariance structure in the repeated measures data. For instance, adding research participants as a random effect in a LME model allows investigators to

resolve the issue of independence among repeated measures by controlling for individual variation among participants. Essentially, the inclusion of subject as a random effect in the model assumes that each participant has a unique intercept, or "baseline", for each variable. Linear mixed-effects models also allow for an understanding of how changes in an individual predictor variable, among other co-existing variables, impact the outcome measure. These regression coefficients provide more detailed information about relationships among predictors and outcome variables than Pearson correlation coefficients as the Pearson correlation coefficient simply measures the strength of the linear relationship between each selected pair of variables independent of the others. Additionally, driven by the research questions and the nature of the independent and dependent variables, researchers can build and compare LME models differing in complexity to best summarize findings. Many possibilities regarding appropriate types of models, necessary data transformations to achieve linearity for each variable, and the inclusion of interaction terms as well as random slopes or intercepts can be considered.

Despite the wide acceptance of the LME method and similar approaches for sophisticated multivariate data analysis, researchers do not necessarily take into account the differences between Pearson correlation and LME models for proper statistical modeling and interpretation of their data. The current report of side-by-side comparison was propelled by the successive publication of two recent electrophysiological studies from our lab that respectively used conventional Pearson correlations and the more sophisticated linear mixed-effects regression models. In particular, our first study investigated whether noised-induced trial-by-trial changes in cortical oscillatory rhythms

in the ongoing auditory electroencephalography (EEG) signal could account for the basic evoked response components in the averaged event-related potential (ERP) waveforms for speech stimuli in quiet and noisy listening conditions (Koerner & Zhang, 2015). When the first study was submitted, we were not aware of the importance and relevance of the LME approach to the analysis of our data set. Even though the paper went through two rounds of revisions, the two anonymous peer reviewers did not raise any concerns for the use of Pearson correlation in our analysis. Our second study further examined whether the noise-induced changes in trial-by-trial neural phase locking, as measured by inter-trial phase coherence (ITPC) and spectral EEG power, could predict averaged mismatch negativity (MMN) responses for detecting a consonant change and a vowel change and whether the cortical MMN response itself could predict speech perception in noise at both the syllable and sentence levels (Koerner & Zhang, 2015). In the publication process of the second study, reviewers questioned the validity of the Pearson correlation analysis for the multiple measures for the same speech stimuli from the same group of subjects, which led to a major revision adopting the LME regression analysis. In hindsight, as the trial-by-trial oscillations and the averaged ERPs are different analysis techniques applied to the same EEG signal, it would have been appropriate to choose the LME models to report the statistical results in our first publication.

As these two previous publications in auditory neuroscience reported only correlation results using one statistical approach, a direct comparison of both the Pearson correlation and LME approaches can be helpful to highlight the differences in the statistical results. Although our examples here are exclusively focused on speech

perception research, the informative comparisons of the statistical results are presented as a further development to advocate for proper implementation of statistical modeling and interpretation of multivariate data analysis in future studies of cognitive neuroscience and experimental psychology.

II. Study 1

Koerner and Zhang (2015) aimed to determine whether noise-induced changes in trial-by-trial neural synchrony in delta (0.5–4 Hz), theta (4–8 Hz), and alpha (8–12 Hz) frequency bands in response to the syllable /bu/ in quiet and in speech babble background noise at a -3 dB SNR (signal-to-noise ratio) were predictive of variation in the N1–P2 ERPs across participants.

A. Statistical Methods

In the published data (Koerner & Zhang, 2015), Pearson correlations were used to examine the strength of linear relationships between ITPC and the N1–P2 amplitude and latency measures pooled across the two listening conditions for each participant and frequency band, resulting in 12 correlations. The reported *p*-values were adjusted for multiple comparisons. Prior to this analysis, scatterplots were used to check the linearity of each pair of continuous variables. Separate repeated measures analysis of variance (ANOVA) were also used to examine the effects of background noise on ITPC and N1–P2 latency and amplitude measures. The ITPC values ranged from 0 to 1, where 1 represents perfect synchronization across trials and 0 represents absolutely no synchronization across trials. Resulting *p*-values were adjusted for multiple comparisons.

For the current comparative report, linear mixed-effects models were developed using R (R Core Team, 2014) and the *nlme* package (Pinheiro et al., 2016). Participants were used as a "by-subject" random effect and listening condition (quiet vs. noise) was included as a blocking variable in each linear mixed-effect model. ITPC values at time points associated with the N1 and P2 responses in delta, theta, and alpha frequency bands were included as fixed effects. For each Pearson correlation and linear mixed-effects model, the significance of each variable in predicting behavioral performance was assessed with the significance level at 0.05.

B. Results

Koerner and Zhang (2015) provided detailed results from repeated measures ANOVAs and the Pearson correlations (see replicated Table 7 for summary of correlation coefficients). The repeated measures ANOVA revealed significant noise-induced delays in N1 ($F(1, 10) = 53.71$, $p < 0.001$) and P2 ($F(1, 10) = 22.27$, $p < 0.001$) latency as well as a significant reduction in N1 amplitude ($F(1, 10) = 13.85$, $p < 0.01$). Additionally, the repeated measures ANOVA revealed significant noise-induced reductions in ITPC for N1 in delta ($F(1, 10) = 20.68$, $p < 0.01$), theta ($F(1, 10) = 18.51$, $p < 0.01$), and alpha ($F(1, 10) = 23.45$, $p < 0.001$) frequency bands as well as for P2 in delta ($F(1, 10) = 13.27$, $p < 0.01$), theta ($F(1, 10) = 14.86$, $p < 0.01$), and alpha ($F(1, 10) = 14.57$, $p < 0.001$) frequency bands.

Results from the Pearson correlation tests showed that ITPC was significantly correlated with N1 latency in delta ($r = -0.586$, $p < 0.01$), theta ($r = -0.521$, $p < 0.05$), and alpha ($r = -0.510$, $p < 0.05$) frequency bands. Similarly, significant correlations were

found between ITPC and N1 amplitude in delta ($r = 0.780$, $p < 0.001$), theta ($r = -0.765$, $p < 0.001$), and alpha ($r = -0.720$, $p < 0.001$) frequency bands. Correlational analysis also revealed significant correlations between ITPC and P2 latency in delta ($r = -0.468$, $p < 0.05$), theta ($r = -0.575$, $p < 0.01$), and alpha ($r = -0.586$, $p < 0.01$) frequency bands as well as between ITPC and P2 amplitude in delta ($r = 0.666$, $p < 0.01$), theta ($r = 0.612$, $p < 0.01$), and alpha ($r = 0.599$, $p < 0.01$) frequency bands.

Table 7. Correlation coefficients for relationship between-phase locking values and N1 and P2 latency and amplitude values in response to the CV syllable /bu/ at electrode Cz as reported in Koerner and Zhang (2015).

	N1		P2	
Frequency Band	Latency	Amplitude	Latency	Amplitude
Delta	-0.586**	-0.780***	-0.468*	0.666**
Theta	-0.521*	-0.765***	-0.575**	0.612**
Alpha	-0.510*	-0.720***	-0.586**	0.599**

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Results from the linear mixed-effects models showed that ITPC in the delta frequency band was a significant predictor of N1 ($F(1, 7) = 16.12$, $p < 0.01$) and P2 amplitude ($F(1, 7) = 10.72$, $p < 0.05$) across listening conditions. Neural synchrony in the alpha frequency band was a significant predictor of N1 latency ($F(1, 7) = 12.51$, $p < 0.05$) across listening conditions. Potential interaction effects were statistically nonsignificant when examined in a full LME model and were therefore removed from the report. An examination of regression coefficients allows for an interpretation of how each fixed effect is related to the outcome measure of interest. For example, a one-point decrease in ITPC in the delta frequency band is associated with a 1.05 unit increase in the N1 amplitude (see Table 8 for a summary of F-statistics and correlation coefficients (B)).

The residual plots from each linear mixed-effects model were normally distributed and did not reveal heteroscedasticity or significant trends. Therefore, it is not expected that generalized linear models would provide better results.

Table 8. F-statistics and regression coefficients (β) for each fixed effect from linear mixed-effects regression models for N1–P2 latencies and amplitudes.

Variable	N1 Latency		N1 Amplitude		P2 Latency		P2 Amplitude	
	F	β	F	β	F	β	F	β
Intercept	964.79***	-	155.62***	-	568.62***	-	31.64***	-
Condition	106.88***	-	16.58**	-	31.93***	-	4.13	-
Delta	0.06	−0.30	16.12**	−1.05	0.46	0.48	10.72*	0.96
Theta	0.46	−0.45	0.17	−1.82	4.01	−0.23	0.00	−0.11
Alpha	12.51**	0.80	3.24	2.01	0.68	−0.41	0.00	0.09

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

III. Study 2

Koerner et al. (2016) aimed to examine whether noise-induced changes in the MMN and spectral power in the theta frequency band in response to a consonant change (/ba/ to /da/) and vowel change (/ba/ to /bu/) in a double-oddball paradigm were predictive of speech perception in noise at the syllable and sentence levels.

A. Statistical Methods

For a direct comparison, Pearson correlations were used to examine correlations between the objective MMN (latency, amplitude, and EEG theta power) in response to /da/ and /bu/ and behavioral responses (percent correct phoneme detection, reaction time, and percent correct sentence recognition) pooled across quiet and speech babble noise listening conditions, resulting in 18 correlations. A check of linearity was performed on each pair of continuous variables using scatterplots. Final *p*-values for each correlation coefficient were adjusted to account for multiple comparisons. As reported in Koerner et

al. (2016), repeated measures ANOVAs were used to examine the effects of background noise on MMN latency, amplitude, and EEG theta power. Linear mixed-effects models were developed to determine whether these objective neural measures were able to predict behavioral performance. Participant was included as a "by-subject" random effect in each linear mixed-effect model while listening condition (quiet vs. noise) and stimulus (/da/ vs. /bu/) were included as blocking (or grouping) variables in each linear mixed-effect model. MMN latency, amplitude, and theta power were added as fixed effects in models with percent correct phoneme detection or reaction time as outcome variables. Similar models were developed to examine whether MMN latency, amplitude, and theta power in response to /da/ or /bu/ were able to predict sentence-level perception using listening condition as a blocking variable. Data transformations for the linear mixed-effects models included re-scaling the MMN latency and behavioral reaction times for phoneme detection as well as log-transforming the percent correct phoneme detection and sentence recognition scores to account for skewness in the data. The significance of each correlation coefficient from the Pearson correlation analysis as well as each fixed effect from the linear mixed-effects models for predicting each behavioral outcome measure was assessed at $\alpha = 0.05$.

B. Results

In the Pearson tests, significant correlations were found between MMN latency recorded in response to the vowel-change and percent correct phoneme detection ($r = 0.53$, $p < 0.05$) for /bu/ as well as percent correct sentence recognition ($r = -0.40$, $p < 0.05$) across the quiet and noise listening conditions. Significant correlations were also

found between MMN amplitude recorded in response to the vowel-change and percent correct phoneme detection ($r = -0.50$, $p < 0.05$) and reaction time ($r = 0.56$, $p < 0.01$) for /bu/, as well as percent correct sentence recognition ($r = -0.66$, $p < 0.01$) across listening conditions. Similar trends were found between theta power in response to the vowel-change and percent correct phoneme detection ($r = 0.41$, $p < 0.05$) and behavioral reaction time ($r = -0.49$, $p < 0.05$) in response to the CV syllable /bu/, as well as behavioral sentence recognition ($r = 0.59$, $p < 0.01$) across listening conditions. Additionally, results revealed significant correlations between MMN latency recorded in response to the consonant-change and percent correct phoneme detection ($r = -0.47$, $p < 0.05$) for /da/ as well as sentence recognition ($r = -0.53$, $p < 0.01$) across the quiet and noise listening conditions (see Table 9 for a summary of correlation coefficients).

Repeated measures ANOVA results from Koerner et al. [55] showed significant effects of background noise on MMN latency ($F(1, 14) = 29.43$, $p < 0.001$), amplitude ($F(1, 14) = 32.52$, $p < 0.001$), and EEG theta power ($F(1, 14) = 19.37$, $p < 0.001$). Koerner et al. [55] also provided detailed results from the linear mixed-effects regression analysis (see replicated Table 10 for summary of regression model results). Linear mixed-effects models showed that both MMN latency ($F(1, 40) = 7.86$, $p < 0.01$) and spectral power in the theta band ($F(1, 40) = 6.61$, $p < 0.05$) were significant predictors of percent correct phoneme detection across listening conditions and stimuli. Additionally, MMN amplitude in response to the syllable /bu/ was a significant predictor of sentence recognition across listening conditions ($F(1, 11) = 7.21$, $p < 0.05$). As all residual plots from each linear mixed-effects model revealed that residuals were normally distributed

without any signs of heteroscedastic variance or significant trends, we do not expect that generalized linear models would improve the results. Interactions were tested in previous models and were subsequently removed due to a lack of statistical significance.

Table 9. Correlation coefficients for brain-behavior correlations between neural MMN latency, amplitude, and theta power for /bu/ and /da/ at electrode Cz and behavioral phoneme detection percent correct, reaction time, and percent correct sentence recognition scores.

	Latency (ms)		**Amplitude (μV)**		**Power (dB)**	
	/bu/	**/da/**	**/bu/**	**/da/**	**/bu/**	**/da/**
Phoneme Detection (% Correct)	−0.53*	−0.47*	−0.50*	−0.17	0.41*	0.13
Reaction Time (ms)	0.34	0.39	0.56**	0.02	−0.49*	0.01
Sentence Recognition (% Correct)	−0.40*	−0.53**	−0.66**	−0.07	0.59**	0.18

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 10. F-statistics and regression coefficients (β) for fixed effects from linear mixed-effects regression models for each behavioral measure (Koerner et al. (2016)).

Variable	**Percent Correct Phoneme Detection**		**Phoneme Detection Reaction Time**		**Percent Correct Sentence Recognition (/bu/)**		**Percent Correct Sentence Recognition (/da/)**	
	F	**β**	**F**	**β**	**F**	**β**	**F**	**β**
Intercept	161.51***	-	4199.98***	-	431.41***	-	335.12***	-
Condition	131.68***	-	61.92***	-	291.32***	-	247.69***	-
Stimulus	114.20***	-	21.05***	-	-	-	-	-
Latency	7.86**	0.61	0.000	0.03	1.24	−0.19	0.44	−0.21
Amplitude	3.10	-0.09	0.002	0.02	7.21*	0.24	0.41	0.05
Theta Power	6.61*	0.05	0.368	0.01	0.46	−0.01	1.50	−0.02

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

IV. Discussion

This current report compared results from Pearson correlations and linear mixed-effects regression models using data from two published ERP studies. It was determined that Pearson correlations were not appropriate for examining relationships in our data,

which contained built-in differences across within-subject repeated measures. The results showed how linear mixed-effects regression models (after verification of normality of residuals and homogeneity of variance) are able to depict relationships between the predictor and outcome variables while taking into account repeated measures across participants. While the LME models were able to confirm basic conclusions gained from the Pearson correlation analyses for both studies (Koerner & Zhang, 2015; Koerner et al., 2016), a comparison of methods and results for each model highlighted differences between the two approaches.

The repeated measures ANOVA indicated that background noise had a significant effect on N1 and P2 latencies as well as N1 amplitudes in response to the syllable /bu/ (Koerner & Zhang, 2015). Similarly, the repeated measures ANOVA revealed that MMN latency, amplitude, and spectral power were significantly impacted by background noise (Koerner et al., 2016). These results support the possibility that pooling data from quiet and noise listening conditions created a built-in contrast and bias between data points when Pearson correlations were used, which partly led to the overestimation of the association strength in the reported results (Tables 1 and 3). In other words, the Pearson correlation analysis ignores these built-in differences and treats this type of data as if each variable in the repeated measures design were independent and normally distributed across the two listening conditions. The resulting *p*-values represent the probability of observing an effect that is as large, or larger, than what would be observed if there was no covariance structure in the repeated measures. In contrast, LME regression analysis was able to account for the covariance structure and grouping factors for the repeated

measures. Tests of significance from the LME models examined whether each predictor variable, or fixed effect, was significantly different than zero while taking into account the other fixed or random effects in the model.

One issue common to regression analysis concerns the possible existence of multi-collinearity (or the existence of high correlations) among the predictor variables and how it may inflate the results with unstable estimates of regression coefficients such as an overall significant model with no significant predictors (McElreath, 2016; Baayen et al., 2008; Bagiella et al., 2000; Magezi, 2015). In the mixed-effects (or multilevel) models, the implementation of fixed and random effects allows control of the within-subject factor for repeated measures, and the additional stepwise approach allows removal of predictor variables in a systematic fashion, for instance, calculating a variance inflation factor (VIF) to identify collinear predictors to aid the stepwise removal of predictors from the LME models. The VIF represents the proportion of variance in one predictor variable accounted for by all the other predictors in the model. Estimation of VIFs for each predictor and progressive dropping of the predictor with the largest VIF beyond the cutoff criterion can be helpful in dealing with the collinearity of interaction terms. By contrast, Pearson correlation analysis assumes independence of the variables, and only fixed effects are directly examined piecewise without elaborate procedures to take into account how the existing associations/differences among the predictor variables may contribute to (oftentimes inflate) the correlation coefficients. The bivariate Pearson correlation analysis disregards potential correlations and data groupings among variables, which makes it inappropriate for research questions that aim to examine associations

between variables that contain built-in differences between experimental conditions or subject groups.

Although the flexibility in model selection can be considered a strength of LME regression analysis, the number of educated choices a researcher must make while developing and implementing models can be a challenge. For instance, the inclusion of interactions or random effects in LME models affects the regression coefficients and interpretation of fixed effects, which cannot properly be taken into account in the bivariate Pearson correlation analysis. Although stepwise regression methods are available as a systematic approach to choose an appropriate model, it is important for researchers to think deeply about the subject matter in order to determine whether the inclusion and interpretation of specific fixed and random effects are appropriate for the specific research question and study objective.

While the two ERP studies reported here are clearly limited in scope and depth of analysis, the side-by-side comparisons clearly demonstrate the limitations and inappropriateness of the Pearson approach as well as its inflated correlation estimation results for the data sets. Given that multiple analysis techniques (for example, waveform analysis, source localization, time-frequency analysis) can be applied to the same neurophysiological data in cognitive neuroscience research (Koerner, Zhang, Nelson, Wang, & Zou, 2016; Koerner & Zhang, 2015; Zhang et al., 2011, 2016), a cautionary note against the convenient use of the simple Pearson correlation test is necessary when selecting and applying statistical models to interpret brain-behavior correlations (e.g.,

biomarkers of various diseases and disorders) or correlations among the various brain measures with prior distributions and covariance structure for repeated measures.

V. Conclusions

In sum, this statistical report compared conventional Pearson correlations and linear mixed-effects (LME) regression models using data from two published auditory electrophysiology studies. The Pearson correlation test is inappropriate for the specific research questions in both studies as the neural responses across listening conditions were simply treated as independent measures. Although our comparative analysis is limited in its scope and depth, this technical note demonstrates the advantages as well as the necessity to apply mixed-effects models to properly account for the built-in relationships among the multiple predictor variables, which has important implications for proper modeling and interpretation of human behavior in terms of neural correlates and biomarkers.

Chapter 5: The P3 as a Neural Correlate of Speech Perception in Noise

Sections I-V are reprinted from:
Koerner, T.K., Zhang, Y., Nelson, P., Wang, B., & Zou, H. (in press). Neural indices of phonemic discrimination and sentence level speech intelligibility in quiet and noise: A P300 study. *Hearing Research.*

I. Introduction

Measures of brain electrical activity have been important in investigating mechanisms that allow listeners to extract target signals from interfering background noise for successful speech communication. Previous auditory event-related potential (AERP) studies have demonstrated the detrimental effects of background noise on the timing and strength of neural responses to speech and non-speech stimuli (Billings et al., 2011; Koerner and Zhang, 2015; Parbery-Clark et al., 2011; Russo et al., 2009). Furthermore, the noise-induced changes in different AERP components have been shown to predict behavioral measures of perceptual and cognitive abilities (Anderson, Parbery-Clark, Yi, & Kraus, 2011; Anderson, Skoe, Chandrasekaran, Zecker, & Kraus, 2010; Billings, Mcmillan, Penman, & Gille, 2013; Billings, Penman, Mcmillan, & Ellis, 2015; Koerner et al., 2016; Song, Skoe, Banai, & Kraus, 2011). The present AERP study represents a sequel to our previous work (Koerner et al., 2016) to determine neural correlates of speech-in-noise perception at the syllable and sentence levels.

Of particular interest to the current report is the auditory P3 response, which is thought to be sensitive to attentional and cognitive processes involved in auditory and speech perception (Picton, 1992; Polich, 2004). The P3 is typically elicited using an active-listening oddball paradigm, in which the subjects are instructed to respond when they detect an infrequent deviant stimulus (e.g., /ba/) within a string of repeated standard

stimuli (e.g., /da/). Several studies have examined the effects of noise on the P3 response for segmental speech processing, including phonemic contrasts such as /ba/ vs /da/ and how various signal-to-noise ratios (SNRs) impacted the P3 response for speech discrimination (Kaplan-Neeman, Kishon-Rabin, Henkin, & Muchnik, 2006; Whiting, Martin, & Stapells, 1998). Two general findings emerged from these studies. First, listening in noise resulted in significant increases in P3 latency and reductions in P3 amplitude for behaviorally discriminable speech stimuli, which were accompanied with increases in behavioral reaction time and reductions in accuracy. Second, Pearson correlation tests showed a negative correlation between sentence intelligibility scores and the P3 peak latency measures for detecting the phonemic (/ba/ vs. /da/) change as well as a positive correlation between behavioral reaction time measures of phonemic discrimination and the corresponding P3 latencies for the phonemic contrast (Bennett et al., 2012). Thus the P3 response appears to be a potential neurophysiological marker for speech-in-noise perception at both segmental and sentential levels. However, there was a limitation in the experimental design of the previous studies as they all focused on consonantal change detection alone. It remains unclear how the P3 responses to consonantal changes differ from those to vowel changes in terms of their predictive power for the behavioral outcome of speech-in-noise perception.

Our work was motivated by the fact that vowels and consonants in a spoken language elicit different patterns of behavioral and neural responses. At the subcortical level, frequency following response (FFR) measures showed larger noise-induced effects on consonant encoding than vowel encoding in the CV syllable context (Anderson, Skoe,

Chandrasekaran, & Kraus, 2010; Anderson, Skoe, Chandrasekaran, Zecker, et al., 2010; Russo et al., 2004; Song, Skoe, Banai, & Kraus, 2011), which is consistent with the behavioral finding of larger noise-induced impact on the detection of consonants compared to vowels (Korczak & Stapells, 2010; Miller & Nicely, 1955; Parikh & Loizou, 2005; Phatak & Allen, 2007; Pickett, 1957). At the cortical level, however, recent studies showed that the neural coding of vowel contrasts as reflected in the mismatch negativity (MMN) response was more susceptible to the presence of background noise compared to consonants (Koerner et al., 2016; Niemitalo-Haapola et al., 2015). Unlike the P3 response that requires an overt behavioral response for detecting a stimulus change and peaks at a later time point than the MMN, the MMN is thought to index pre-attentive automatic change detection independent of focused attention (Näätänen, Paavilainen, Rinne, & Alho, 2007). In our previous MMN study (Koerner et al., 2016), we argued that the differential effects of noise on the MMNs for consonant discrimination and vowel discrimination reflected their different contributions to speech intelligibility in noise as evidenced in multiple behavioral studies (Cole et al., 1996; Fogerty and Humes, 2012, 2010; Fogerty and Kewley-Port, 2009; Kewley-Port et al., 2007; Owren and Cardillo, 2006). On this point, there is an abundance of neuroimaging data indicating that distinct brain mechanisms are involved in processing consonants and vowels (Caramazza, Papagno, & Ruml, 2000; Carreiras & Price, 2008; Carreiras, Vergara, & Perea, 2009).

To date, previous speech-in-noise research studies have not directly compared cortical processing of consonant discrimination and vowel discrimination in the presence or absence of background noise using the attention-driven P3 response. The current P3

study followed up our previous MMN work with a double-oddball paradigm (Koerner et al., 2016) in which two deviants consisting of either a consonant change (from /ba/ to /da/) or a vowel change (from /ba/ to /bu/) are presented within the same listening session. We investigated the role of attention in the neural processing of consonant and vowel stimuli by examining the differential effects of background noise on neural responses as listeners actively detected each of these stimulus contrasts. We were particularly interested in examining how the segmental-level P3 responses for consonant and vowel stimuli jointly or separately contribute to sentence-level performance.

In addition to conventional analysis on the latency and amplitude of AERP components, researchers have also begun to use time-frequency analysis techniques to determine how experimental stimulus and task factors impact induced and evoked cortical oscillations within the ongoing EEG signal. The oscillations are thought to play a key role in enabling sensory and cognitive processing across and within cortical networks (Başar, Demiralp, Schürmann, Başar-Eroglu, & Ademoglu, 1999; Klimesch, Sauseng, Hanslmayr, Gruber, & Freunberger, 2007; Koerner & Zhang, 2015; Makeig, Debener, Onton, & Delorme, 2004; Sauseng et al., 2007; Zhang et al., 2011). Specifically, oscillations in the delta (1-4 Hz), theta (4-8 Hz), and alpha (8-12 Hz) frequency bands have been found to be associated with the cortical P3 response, which may represent underlying cognitive demands related to different processes of signal processing and attentional engagement (Demiralp, Ademoglu, Istefanopulos, Basar-Eroglu, & Basar, 2001; Intriligator & Polich, 1994, 1995; Kolev, Demiralp, Yordanova, Ademoglu, & Isoglu-Alkac, 1997; Polich, 1997; Spencer & Polich, 1999; Yordanova &

Kolev, 1998). For example, Basar-Eroglu and colleagues (1992) examined the functional significance of delta and theta oscillations using two auditory P3 paradigms: an omitted stimulus paradigm, which required attention to every third signal, and an oddball paradigm, which requires additional signal matching and decision making processes to respond to rare, randomly presented target stimuli. From their results, it was suggested that theta oscillations, which were altered across both tasks, are functionally related to focused attention and signal detection processes. On the other hand, delta band activity only increased during the oddball paradigm, suggesting delta oscillations are more specifically related to signal matching and decision making processes. Furthermore, event-related oscillations in the alpha band have been shown to reflect top-down processing resources that are important for inhibiting task irrelevant maskers, such as during speech perception in complex listening environments (Klimesch, 2012; Straub, Wostmann, & Obleser, 2014; Wilsch, Henry, Herrmann, Maess, & Obleser, 2014). Together, the existing literature shows the importance of time-frequency analysis to specify delta, theta, and alpha oscillatory activities underlying the auditory P3 responses for consonant discrimination and vowel discrimination that are differentially affected by the background noise. For the current study, we will examine event-related cortical oscillations for the P3 responses in terms of neural synchrony and EEG spectral power across trials.

We hypothesized that background noise would significantly impact the P3 response as well as oscillatory activities in delta, theta, and alpha frequency bands in response to both consonant and vowel contrasts. Based on the behavioral data from our

previous study (Koerner et al, 2016), we expected that background noise would have a larger effect on the detection of the consonant change compared to the vowel change in the attentive listening condition. Furthermore, we predicted that some of the AERP and time-frequency measures would be significant predictors of behavioral performance at both the segmental and sentence levels. While we could not formulate exact hypotheses about the exact brain-behavioral correlates involving both ERP waveform and time-frequency measures, the exploratory results from the current study with an active listening condition would complement those from our previous MMN work that did not require focused attention to detect the consonant and vowel contrasts in the double-oddball paradigm (Koerner et al., 2016).

II. Methods

A. Subjects

The participants were 16 individuals (mean age = 22.5 years, age range = 19 – 32 years, 6 males, 10 females) with normal hearing (as shown in standard audiological assessment with hearing thresholds < 25 dB HL for pure tones from 0.25 to 8 kHz). All participants were right handed native speakers of American English, and had no history for speech, language, or cognitive disorders or difficulties. The Human Research Protection Program at the University of Minnesota approved the research protocol, and all participants provided informed consent prior to beginning the study.

B. Stimuli

Three consonant-vowel (CV) syllables, /ba/, /da/, and /bu/, were synthesized using a 10 kHz sampling rate in the HLsyn software program (Sensimetrics Corporation, USA)

(Koerner & Zhang, 2015). Each syllable was 170 ms in duration with a steady fundamental frequency of 100 Hz and a steady F4 of 3300 Hz. For the /ba/ sound, the HLsyn software generated formant transitions in the first 50 ms of the CV syllables with onset frequencies at 328 Hz, 1071 Hz, and 2298 Hz respectively for F1, F2, and F3. The F1, F2, and F3 onset frequencies were set at 362 Hz, 1832 Hz, and 2540 Hz for /da/ and 230 Hz, 900 Hz, and 2480 Hz for /bu/ respectively. For the vowel portion (50-170 ms) of the /ba/ and /da/ syllables, the steady center F1, F2, and F3 frequencies were 674 Hz, 1140 Hz, and 2350 Hz. The steady center F1, F2, and F3 frequencies were 320 Hz, 860 Hz, and 2620 Hz for the vowel portion of /bu/. IEEE sentences (IEEE, 1969) were also used to obtain sentence recognition scores. We used a four-talker speech babble background noise that was adopted from the Quick Speech In Noise Test (Quick-SIN) (Niquette et al., 2001). The speech and noise stimuli were resampled at 44.1 kHz and were normalized to create a -3 dB SNR using Sony SoundForge 9.0 (Sony Creative Software, USA).

C. Procedure

The reported EEG data were taken from a larger scale study with a two-hour recording session including both passive and active listening conditions (Koerner, Zhang, & Nelson, 2013). All ERP and behavioral test sessions were conducted in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems). The CV syllables were presented via bilateral Etymotic ER-2 insert headphones using EEvoke software (ANT Inc., Netherlands). The speech signals were presented at 60 dB SL relative to the individual listeners' hearing threshold at 1 kHz (Koerner et al., 2016; Koerner & Zhang,

2015; Nie, Zhang, & Nelson, 2014; Rao, Rishiq, Yu, Zhang, & Abrams, 2017). Participants were presented with two listening conditions in both the EEG and behavioral tests: signals in quiet and signals in a four-talker speech babble noise.

As described in our previous MMN study (Koerner et al., 2016), the double-oddball paradigm included two speech contrasts representing a vowel change (from /ba/ to /bu/) and a consonant change (from /ba/ to /da/) within the same recording session. Unlike the passive listening MMN protocol, the listeners in the current P3 study were asked to press a keyboard response button each time they heard a deviant stimulus representing either a consonant or vowel change. The back vowels /a/ and /u/ differ primarily in the steady vowel F1 while the /b/ and /d/ consonants represent transient differences in place of articulation and are cued by second and third formant frequency transitions, which are in a frequency range commonly affected by hearing impairment (Ladefoged, 2006; Miller & Nicely, 1955). The order of stimulus presentation was pseudo-randomized in the double-oddball paradigm so that no blocks began with a deviant stimulus and no two deviants were presented in succession. The standard stimulus /ba/ had a probability of occurrence of 0.75 and both /da/ and /bu/ each had a probability of occurrence of 0.125. Each listening condition consisted of 10 blocks for a total of 832 trials for standard stimuli and 104 trials for each deviant stimulus. Participants were given 10-15 sec breaks between blocks. The interstimulus interval (ISI) was 1000 ms with a 100 ms randomization from trial to trial. The quiet and noise listening conditions were counter-balanced across subjects to reduce potential test order

effects. During the recording session, participants were instructed to relax and minimize excessive movements.

Sentence recognition performance was recorded in an additional 30-minute behavioral session using randomized lists of IEEE sentences presented through TDH-39 headphones. The IEEE lists consisted of 10 low context sentences with 5 key words in each sentence. For both the quiet and noise conditions, two sentence lists of 50 key words were presented from one female and one male talker for a total of 100 key words per listening condition (IEEE, 1969). Participants were instructed to repeat the IEEE sentences they heard out loud, as best as they could, while a certified audiologist evaluated the word-by-word responses for recognition accuracy.

D. Data Analysis

ERP measures

Continuous EEG data were recorded using the Advanced Neuro Technology EEG System (Advanced Source Analysis version 4.7) and a 64-channel Ag AgCl electrode WaveGuard cap with a REFA-72 amplifier (TMS International BV) (bandwidth = 0.016-200 Hz, sampling rate = 512 Hz). The average impedance of electrodes was below 5 kOhms. ERP waveform analysis was completed offline in BESA (Version 6.0, MEGIS Software GmbH, Germany). The EEG data were bandpassed at 0.5-40 Hz. The ERP epoch length consisted of a 100 ms prestimulus baseline and a 700 ms poststimulus interval. Automatic artifact rejection criteria were set at ± 50 μV. The P3 was analyzed with an averaged mastoid reference at the Pz electrode. The time window for assessing P3 peak latency was 250-680 ms, which was based on the grand average waveforms in

the quiet and noise conditions relative to the pre-stimulus interval. Computation of the P3 amplitude used an integration window of 40 ms centered at peak. Similar time windows for P3 quantification were used in our previous P3 studies (Nie et al, 2014; Rao et al., 2017).

Two event-related time frequency analysis measures were computed to evaluate trial-by-trial cortical oscillations in delta, theta, and alpha frequency bands: inter-trial phase coherence (ITPC) and event-related spectral perturbation (ERSP). Inter-trial phase coherence evaluates the synchronization of trial-by-trial oscillations as a function of time and frequency (David, Kilner, & Friston, 2006; Makeig et al., 2004; Nash-Kille & Sharma, 2014): $ITPC_{(t,f)} = \frac{1}{n}\sum_{k=1}^{n}\frac{F_k(t,f)}{|F_k(t,f)|}$, where F stands for the Fourier transform, t stands for time, f is frequency, k is the trial number, n is the total number of accepted trials, and | | is the complex norm (Delorme & Makeig, 2004). Resulting ITPC values in a given frequency band can range from 0, which represents no synchronization across trials, to 1, which represents perfect synchronization across trials. Event-related spectral perturbation is used to examine trial-by-trial change in evoked power (in dB) from pre-stimulus baseline as a function of frequency and time (Fuentemilla, Marco-Pallarés, & Grau, 2006; Makeig, 1993): $ERSP_{(f,t)} = \frac{1}{n}\sum_{k=1}^{n}|F_k(f,t)|^2$, where F stands for the Fourier transform, t is time, f is frequency, k is the trial number, and n is the total number of accepted trials, (Delorme & Makeig, 2004).

Analysis of ITPC and ERSP at electrode Pz was completed using the *newtimef* function in EEGLAB (Delorme & Makeig, 2004). A short-term Fourier Transform (STFT) with Hanning window tapering (Koerner & Zhang, 2015; Koerner et al., 2016)

was adopted to extract the ITPC and ERSP values for the delta, theta, and alpha frequency bands, which is recommended for the analysis of low frequency activities. In order to overcome restrictions from the use of fixed windows in conventional analysis, the modified STFT method used overlapping sliding windows adapted to the target frequency bins and zero-padding was applied for short epochs that did not have sufficient sample points for the Fourier transform. Estimated frequencies were from 0.5-40 Hz with a step interval of 0.5 Hz. The analysis time windows for both deviant stimuli were 300-400 ms for the quiet listening condition and 400-500 ms for the listening condition with background noise. The ITPC and ERSP values represented the maximum in the defined time windows, which were chosen based on peak latency data from individual P3 waveforms.

The statistical analyses from both the AERP and behavioral portions of the study were analyzed in R (R Core Team, 2014). A repeated measures analysis of variance (RM-ANOVA), with $\alpha = 0.05$, was conducted to examine the statistical significance of stimulus type (/da/ or /bu/), listening condition (quiet or noise), and any potential interactions (stimulus x condition) on P3 latency, amplitude, ITPC, and ERSP at electrode Pz. Where significant interaction effects were observed, tests of simple main effects and post hoc two-tailed t-tests for selected factors of interest were also conducted to evaluate how consonant- and vowel- changes were processed differently in quiet and in noise.

Behavioral Measures

Percent correct change discrimination scores and reaction time for the detection of consonant- and vowel-changes in the double-oddball paradigm were obtained from the button-press responses recorded during the quiet and noise conditions. A RM-ANOVA was completed to investigate the statistical significance of stimulus type (/da/ or /bu/), listening condition (quiet or noise), and any potential interactions (stimulus x condition) on behavioral discrimination accuracy and reaction time. A RM-ANOVA was carried out to examine the significance of listening condition (quiet vs. noise) on IEEE sentence recognition.

Analysis of Brain-Behavior Relationships

Linear mixed-effects (LME) regression models were developed in R (R Core Team) using the *nlme* package (Pinheiro et al., 2016) to examine whether neural measures were predictive of behavioral performance at the segmental and sentence levels across participants, listening conditions, and stimuli (Koerner & Zhang, 2017; Koerner et al., 2016). Similar statistical techniques have been used previously to examine links between neurophysiological and behavioral measures of speech perception (Billings et al., 2013; 2015). These models allow for an examination of multiple neural measures as predictor variables, or fixed effects, on a particular outcome variable while taking into account repeated measures across participants (Baayen et al., 2008; Bagiella et al., 2000; Magezi, 2015).

Linear mixed-effects models were created with by-participant intercept as a random effect. Listening condition (quiet vs. noise) and stimulus (/bu/ vs. /da/) were added as blocking variables. P3 latency and amplitude as well as ITPC and ERSP in

delta, theta, and alpha bands were added as fixed effects. Outcome variables included phoneme change detection performance, reaction time, and sentence-level recognition. Final reduced models with fewer terms were chosen with the Akaike's Information Criterion (AIC) using stepwise linear regression with bidirectional elimination on main effects. The models used to examine percent correct sentence recognition did not contain stimulus type as a fixed effect because sentence recognition performance only varied by listening condition. Data transformations included re-scaling P3 latency and behavioral reaction time values as well as log-transforming percent correct phoneme detection and sentence recognition scores to account for skewness in the data. For each LME model used in this study, the significance of each variable in predicting behavioral performance was assessed with $\alpha = 0.05$.

III. Results

Analysis of phoneme- and sentence-level behavioral data (Table 11) showed that background noise had a significant impact on percent correct phoneme change detection and reaction time as well as percent correct sentence recognition. Background noise also impacted P3 latency and amplitude (Fig. 7, Table 12) as well as trial-by-trial neural synchrony (Fig. 8, Table 13). Several significant neural markers of behavioral speech perception at both the phoneme- and sentence-levels were revealed using stepwise linear-mixed effects regression models (Table 14).

A. Percent Correct Phoneme Discrimination and Sentence Recognition

As expected, RM-ANOVA revealed a significant detrimental effect of background noise on behavioral reaction time ($F(1,15) = 47.12$, $p < 0.001$) and percent

correct phoneme detection ($F(1,15) = 94.41$, $p = < 0.001$) (Table 11). More specifically, behavioral reaction time was significantly prolonged and percent correct phoneme detection was significantly reduced in background noise compared to the quiet listening condition. RM-ANOVA also revealed a significant effect of stimulus on behavioral reaction time ($F(1,15) = 28.77$, $p < 0.001$) and percent correct phoneme detection ($F(1,15) = 147.40$, $p < 0.001$), such that reaction time was significantly longer and percent correct phoneme detection significantly poorer during behavioral discrimination of the consonant change (from /ba/ to /da/) compared to the vowel change (from /ba/ to /bu/). There was also a significant effect of background noise on sentence-level recognition ($F(1,15) = 83.44$, $p < 0.001$).

Additionally, RM-ANOVA revealed a significant interaction between listening condition and stimulus for percent correct phoneme detection ($F(1,15) = 20.94$, $p < 0.001$). Post-hoc t-tests indicated that the presence of background noise significantly decreased percent correct detection of the consonant change ($t(1,15) = 16.35$, $p < 0.001$) and had a smaller, yet still significant effect on detection of the vowel change ($t(1,15) = 3.61$, $p < 0.01$).

Table 11. Means (standard error) for behavioral percent correct detection of vowel (/bu/) and consonant (/da/) changes, behavioral reaction time for vowel (/bu/) and consonant (/da/) changes, as well as percent correct behavioral sentence recognition performance.

	Listening Condition	
Behavioral Measure	Quiet	Noise
/bu/ detection (%)	97.77 (0.71)	91.35 (2.12)
/da/ detection (%)	91.57 (1.89)	33.35 (3.84)
Sentence recognition (%)	99.31 (0.35)	67.25 (3.53)
/bu/ response time (ms)	468.94 (9.41)	553.06 (15.82)
/da/ response time (ms)	513.01 (11.97)	622.91 (11.36)

B. Averaged P3 Amplitude and Latency Measures

As predicted, RM-ANOVA revealed significant effects of background noise on P3 latency ($F(1,15) = 15.85$, $p < 0.01$) and amplitude ($F(1,15) = 23.5$, $p < 0.001$) (Table 12 and Figure 7). The neural responses to the two deviant speech stimuli in quiet tended to have shorter latencies and larger amplitudes than responses in background noise. There was also a significant effect of stimulus on P3 amplitude across the quiet and noise listening conditions ($F(1,15) = 21.7$, $p < 0.001$), such that P3 amplitude was smaller in response to /da/ compared to /bu/. However, there was no significant difference in P3 latency between responses to /da/ and /bu/ ($F(1,15) = 2.60$, $p = 0.128$) across listening conditions. Additionally, there were no significant interactions between listening condition and stimulus for P3 latency ($F(1,15) = 0.22$, $p = 0.64$) or amplitude ($F(1,15) = 1.48$, $p = 0.24$).

Table 12. Mean (standard error) values for P3 latency (ms) and amplitude (μV) in response to the CV syllable vowel change (/bu/) and consonant change (/da/) at electrode Pz in quiet and in noise.

	Latency (ms)		Amplitude (μV)	
	Quiet	Noise	Quiet	Noise
/bu/	362.16 (27.34)	459.06 (20.6)	7.64 (0.69)	6.29 (0.77)
/da/	399.29 (27.33)	480.54 (31.61)	5.97 (0.6)	3.31 (0.41)

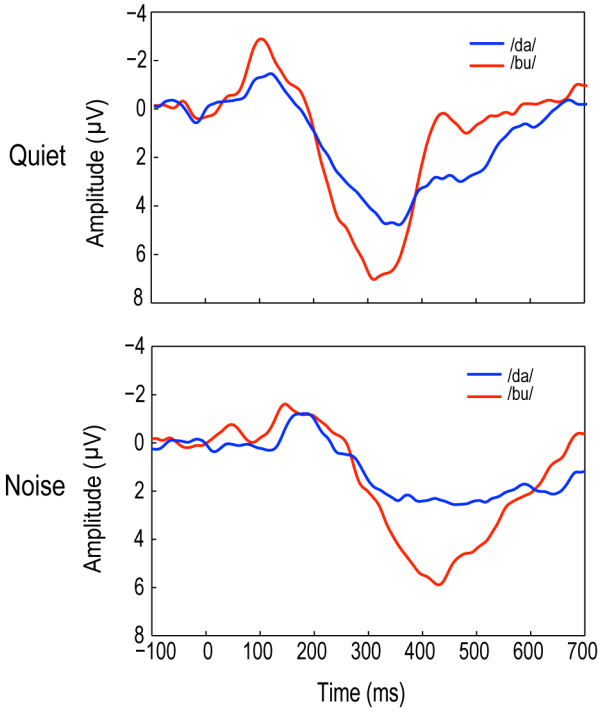

Figure 7. Grand mean P3 waveforms at the Pz electrode depicting responses averaged across participants to the consonant-vowel syllable /bu/ (red) and /da/ (blue) in the quiet (top row) and speech-babble (bottom row) listening conditions. An analysis window of 250-680 ms was used to extract individual P3 latencies.

C. ITPC and ERSP

RM-ANOVA indicated a significant effect of noise on ITPC in the delta ($F(1,15) = 7.68$, $p < 0.05$), theta ($F(1,15) = 12.02$, $p < 0.01$), and alpha ($F(1,15) = 11.03$, $p < 0.01$) frequency bands across the two deviant syllables, /da/ and /bu/ (Table 13 and Figure 8). In contrast, there was no significant effect of background noise on ERSP in the delta

(F(1,15) = 0.11, p = 0.75), theta (F(1,15) = 0.53, p = 0.48), or alpha (F(1,15) = 0.61, p = 0.45) frequency bands due to the existence of very large inter-subject variability of across-trial EEG spectral power for the P3 component relative to the baseline (Table 3 and Figure 3). There were also no significant effects of stimulus on ITPC for delta (F(1,15) = 0.14, p = 0.71), theta (F(1,15) = 0.21, p = 0.65), or alpha (F(1,15) = 0.06, p = 0.81) frequency bands or in ERSP in delta (F(1,15) = 0.51, p = 0.48), theta (F(1,15) = 0.13, p = 0.72), or alpha (F(1,15) = 0.15, p = 0.70) frequency bands.

Table 13. Mean (standard error) ITPC and ERSP values in response to the CV syllable vowel change (/bu/) and consonant change (/da/) at electrode Pz in quiet and in noise in delta, theta, and alpha frequency bands.

ITPC	Delta		Theta		Alpha	
	Quiet	Noise	Quiet	Noise	Quiet	Noise
/bu/	0.48 (0.02)	0.4 (0.02)	0.47 (0.02)	0.38 (0.02)	0.47 (0.02)	0.37 (0.02)
/da/	0.46 (0.02)	0.4 (0.01)	0.44 (0.02)	0.39 (0.02)	0.44 (0.02)	0.39 (0.02)
ERSP						
/bu/	0.63 (0.51)	0.45 (0.52)	0.68 (0.57)	0.39 (0.5)	0.53 (0.53)	0.24 (0.47)
/da/	0.79 (0.49)	0.74 (0.40)	0.73 (0.51)	0.56 (0.51)	0.58 (0.49)	0.41 (0.37)

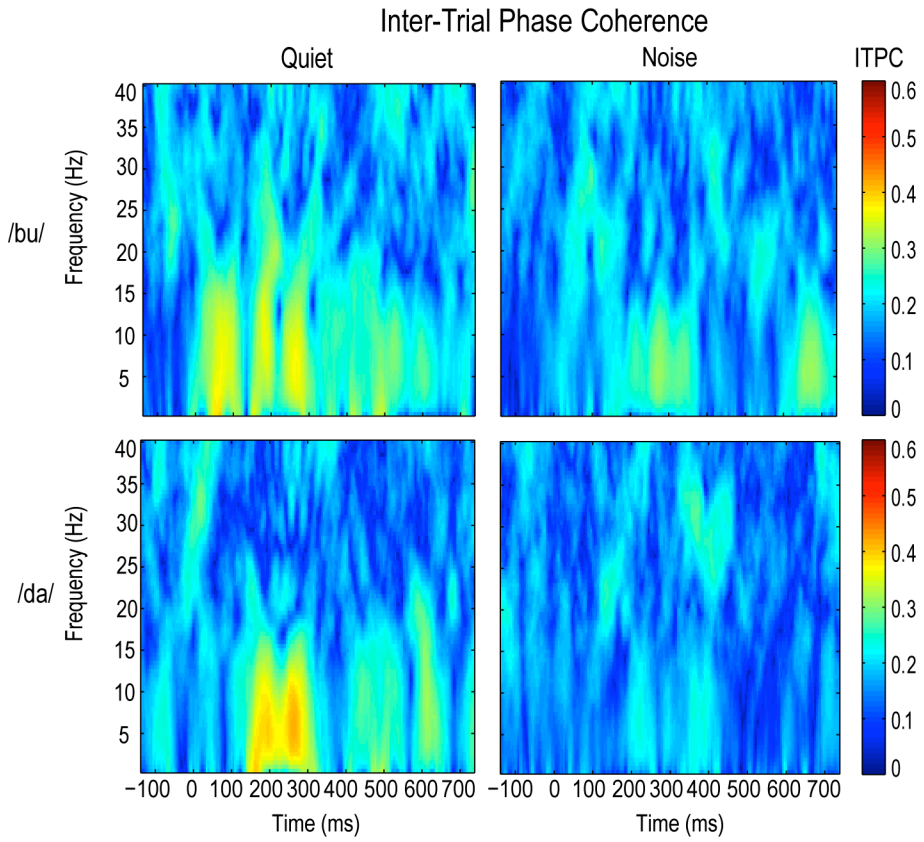

Figure 8. Grand mean inter-trial phase coherence (ITPC) data in both quiet (left column) and babble noise (right column) listening conditions. ITPC values range from 0-1 for change-detection responses to the CV syllables /bu/ (top row) and /da/ (bottom row). ITPC was calculated using an analysis window of 300-400 ms for the quiet listening condition and a window of 400-500 ms for the listening condition with background noise.

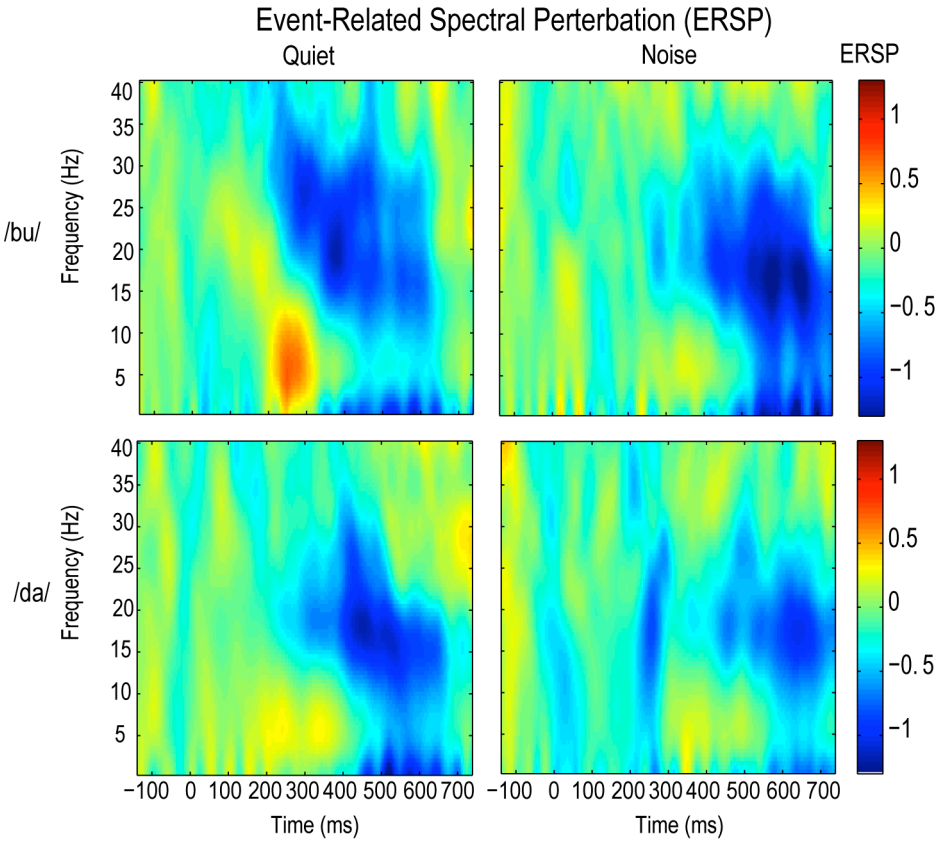

Figure 9. Grand mean event-related spectral perturbation (ERSP) in response to the CV syllables /bu/ (top row) and /da/ (bottom row) in both quiet (left column) and babble noise (right column) listening conditions. ERSP was calculated using an analysis window of 300-400 ms for the quiet listening condition and a window of 400-500 ms for the listening condition with background noise.

D. Linear Mixed-Effects Regression Models

Stepwise linear regression using bidirectional elimination by AIC was used to obtain reduced linear mixed-effects regression models with fewer terms. Table 14 shows a summary of reduced model outputs for segmental-level and sentence-level behavioral measures. The linear mixed-effects regression models used to predict segmental-level performance revealed that P3 amplitude was a significant predictor of percent correct segmental-level change detection ($F(1,42) = 7.01$, $p < 0.05$) and reaction time ($F(1,43) = 11.03$, $p < 0.01$) across listening conditions and stimuli. Additionally, ERSP in the alpha

band was a significant predictor of percent correct segmental-level change detection ($F(1,42) = 8.29$, $p < 0.01$) and reaction time ($F(1,43) = 4.86$, $p < 0.05$) across listening conditions and stimuli. ITPC in the theta band showed a similar but insignificant trend as a predictor of percent correct segmental-level change detection ($F(1,12) = 2.93$, $p < 0.10$). The models used to predict sentence-level performance revealed that ITPC in the theta frequency band ($F(1,12) = 9.69$, $p < 0.01$) in response to /bu/ was a significant predictor of percent correct sentence-level performance across listening conditions. Although ITPC in the alpha band showed a similar trend, it did not reach statistical significance ($F(1,12) = 4.36$, $p < 0.10$) for being a predictor of sentence-level performance. There was also a similar but insignificant trend in ERSP of the alpha frequency band in response to /da/ as a predictor of percent correct sentence-level performance across listening conditions ($F(1,12) = 3.53$, $p < 0.10$). The residual plots from all linear mixed-effects models appeared normally distributed and did not show any signs of heteroscedastic variance or significant trends.

Table 14. F-statistics for fixed effects (P3 latency, amplitude, ITPC and ERSP in delta, theta, and alpha bands) included in final reduced linear mixed-effects regression models for each behavioral measure.

Variable	Percent Correct Phoneme Detection	Phoneme Detection Reaction Time	Percent Correct Sentence Recognition (/bu/)	Percent Correct Sentence Recognition (/da/)
Intercept	269.97***	5439.97***	295.46***	398.83***
Condition	115.72***	92.89***	667.05***	398.74***
Stimulus	104.55***	32.02***	-	-
Latency	-	-	-	-
Amplitude	7.01*	11.03**	-	-
Delta ITPC	-	-	2.06	3.06
Theta ITPC	2.93	-	9.69**	-
Alpha ITPC	-	-	4.36	-
Delta ERSP	-	-	-	1.96
Theta ERSP	0.34	0.87	-	-
Alpha ERSP	8.29**	4.86*	-	3.53

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

IV. Discussion

This ERP study was designed to examine the effects of background noise on P3 latency, amplitude, ITPC, and ERSP in response to a consonant and vowel change and also aimed to determine whether these neural measures are significant neurophysiological predictors of segmental- and sentence-level speech perception across quiet and noise listening conditions.

A. Attention and differential effects of noise on neural coding of consonants and vowels

Our neurophysiological results are consistent with previous reports of noise-induced increases in P3 latency and decreases in P3 amplitude in response to speech (Bennett, Billings, Molis, & Leek, 2012; Martin, Sigal, Kurtzberg, & Stapells, 1997;

Martin & Stapells, 2005; Whiting et al., 1998). Our data add to the literature by revealing a differential effect of background noise on the attentive cortical processing of consonant and vowel stimuli, such that background noise had a larger impact on P3 responses to the consonant change compared to the vowel change, which is consistent with results from our behavioral phoneme-change detection data. However, this pattern of noise-induced effects on the neural coding of consonant and vowel changes appears to be opposite to previous reports from the cortical, pre-attentive MMN response to consonants and vowels in noise (Koerner et al., 2016; Niemitalo-Haapola et al., 2015). Using the same consonant and vowel changes as used in the current study, Koerner et al. (2016) showed that background noise actually had a larger impact on MMN amplitude and EEG theta power in response to the vowel change compared to the consonant change (see direct comparison of MMN and P3 results in Table 15). Koerner et al. (2016) proposed that the larger noise induced effects on the neural processing of the vowel change may have occurred due to the need to internally resolve more transient, aperiodic consonant information in noise when deviant consonant and vowel changes were juxtaposed in a double-oddball paradigm at the pre-attentive processing level.

A main difference in EEG recording protocols between our previous MMN study (Koerner et al., 2016) and the current P3 study is that the P3 response required the participant to pay attention to all stimulus presentations and make an overt behavioral response to deviant stimuli. The opposite patterns regarding the noise impacts on consonants and vowels in our two studies suggest that differences in relative impacts on consonant and vowel change detection results across the two studies are modulated by the

level of focused attention to the target stimulus. The observed differences between passive-MMN results and the active-P3 results may arise due to the fact that they are inherently different components arising from different generators and underlying sensory/cognitive processes. In our pre-attentive MMN paradigm (Koerner et al., 2016), participants were asked to ignore stimulus presentations and focus on a silent, subtitled movie. When attention was directed away from the experimental speech stimuli in this way, the relative importance of the weaker, transient, and aperiodic consonant change may have increased, causing noise to have less of an effect on pre-attentive neural processing of this stimulus compared to the more robust formant differences in the vowel contrast. A study by Gordon, Eberhardt, and Rueckl (1993) provided supporting evidence for our interpretation here. In that study, the relative importance of voice onset time (VOT) and fundamental frequency (F0) cues changed during a /ba/-/pa/ discrimination task in quiet was found to depend on the amount of attention paid to the task. It has been suggested that VOT is the primary cue for discriminating phonemes that differ in voicing while the onset frequency of F0 is considered a weaker acoustic cue. When participants were directly focused on the discrimination task, the stronger VOT cues dominated perception. However, when participants were asked to perform a distractor task during stimulus discrimination, results showed that the relative importance of the stronger VOT cues decreased while the impact of the weaker F0 cues to perception increased (Gordon, Eberhardt, & Rueckl, 1993). Similarly, when participants devoted full attention to the discrimination task in our active P3 paradigm, the weaker consonant contrast that was more difficult to behaviorally discriminate in noise showed weaker

neural responses while responses to the vowel change were less impacted by the presence of background noise.

Table 15. A comparison of mean (standard error) amplitudes from MMN (Koerner et al., 2016) and P3 responses to /bu/ and /da/ in quiet and noise listening conditions. The right column displays results from paired t-tests that examined differences between the quiet and noise listening conditions for each response.

MMN	Quiet	Noise	Paired T-Test (t-statistic)
/bu/	-2.11 (0.3)	-0.40 (0.12)	-6.30***
/da/	-1.16 (0.20)	-1.02 (0.18)	-0.65
P3			
/bu/	7.64 (0.69)	6.29 (0.77)	1.29
/da/	5.97 (0.6)	3.31 (0.41)	3.69**

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

B. Effects of noise on event-related cortical oscillations

Our results revealed that ITPC in delta, theta, and alpha frequency bands in response to both deviant speech stimuli was significantly impacted by the presence of background noise. In contrast, results showed no significant noise-induced effects on ERSP across frequency bands in response to both deviant speech stimuli. On average, it appeared as if ERSP was reduced in noise compared to the quiet listening condition across stimuli; however, there was very large inter-subject variability in the data across listening conditions (see Table 13). These results are consistent with those from Koerner and Zhang (2015), which showed that background noise significantly disrupted averaged trial-by-trial neural synchrony, but did not significantly impact trial-by-trial spectral power (Koerner & Zhang, 2015). Fuentemilla et al. (2006) reported similar results while examining cortical N1 amplitude reduction in response to repeated pure tones; they found that stimulus-evoked phase synchronization occurred without any trial-by-trial spectral

power modulation. These findings appear to support the "phase resetting" model, where stimulus-evoked phase synchronization of cortical oscillations has been shown to at least partially contribute to the neural generation of AERPs (Gruber, Klimesch, Sauseng, & Doppelmayr, 2005; Klimesch et al., 2007; Makeig et al., 2004). In other words, reduced ITPC without noticeable concomitant changes in ERSP across quiet and noise listening conditions supports the possibility that partial stimulus-related phase synchronization of cortical oscillations drives the neural generation of the P3 response to speech.

C. The P3 as a neurophysiological marker of behavior

As described in previous work (Koerner et al. 2016), a -3 dB SNR was chosen because it was shown to result in performance on the speech perception measures used in this study without ceiling or floor effects. We assumed that the cortical P3 response would reflect this variability in behavioral performance based on findings from previous studies that revealed significant relationships between speech-evoked AERPs and behavioral speech perception in background noise (Anderson et al., 2011; Anderson, Skoe, Chandrasekaran, & Kraus, 2010; Anderson, Skoe, Chandrasekaran, Zecker, et al., 2010; Bennett et al., 2012; Billings et al., 2013, 2015; Cunningham et al., 2001; Koerner et al., 2016; Martin & Stapells, 2005; Martin et al., 1999; Muller-Gass, Marcoux, Logan, & Campbell, 2001; Song et al., 2011).

Results from our stepwise linear mixed-effects regression models are consistent with previous studies that have shown significant relationships between the cortical P3 amplitude and speech perception at the segmental-level (Kaplan-Neeman et al., 2006; Martin et al., 1997; Martin & Stapells, 2005; Whiting et al., 1998). During stepwise

regression analyses, P3 latency was consistently removed from our final reduced regression models (see Table 14). Although P3 latency is thought represent the speed of stimulus classification, it has been reported that it is unrelated to behavioral response accuracy or reaction time (McCarthy & Donchin, 1981; Verleger, 1997), which is consistent with our data showing that P3 latency was not able to significantly predict behavioral performance at the segmental- and sentence-levels across listening conditions and stimuli. Unlike results reported by Bennett et al. (2012), our data did not reveal that the averaged cortical P3 latency or amplitude response was able to predict sentence-level recognition across listening conditions or deviant speech stimuli. This could be caused by differences in analysis methods, as Bennett et al. (2012) used Pearson correlations to examine brain-behavior relationships while the current work used stepwise regression models that contain multiple neural measures as fixed effects (Koerner & Zhang, 2017).

Our data also revealed that both ITPC and ERSP in various frequency bands at time points corresponding to the cortical P3 are predictive of segmental- and sentence-level behavioral performance. In other words, impaired speech perception may be indexed by noise-induced reductions in trial-by-trial neural synchrony or power recorded in response to consonant and vowel changes in an active change detection task. Although ITPC in the alpha band in response to the vowel change detection did not reach statistical significance as a predictor of sentence-level performance, ITPC in all three frequency bands in response to the vowel change were the most important for predicting sentence recognition. ERSP in the alpha frequency band for the consonant change detection was also a significant predictor of segmental-level performance across stimuli and listening

conditions as well as sentence-level performance across conditions. These results support the functional associations between alpha oscillations and top-down inhibitory mechanisms (Klimesch, 2012; Straub et al., 2014; Wilsch et al., 2014). In the context of the present study, these selective inhibitory mechanisms may be related to the ability to accurately perceive speech in masking background noise.

D. Novelty and Future Directions

Unlike the previous studies using only one deviant speech stimulus (Kaplan-Neeman et al., 2006; Martin et al., 1997; Martin & Stapells, 2005; Whiting et al., 1998), the present study examined the role of attention in how background noise influences the differential neural processing of consonant and vowel changes using a double-oddball paradigm in relation to behavioral perception. Although researchers have previously linked noise-induced variability in the P3 response with sentence-level perception in noise (Bennett et al., 2012), our results add to this body of knowledge by highlighting the different contributions by consonants and vowels to speech intelligibility in noise at both segmental and sentential levels. Furthermore, our results revealed that important complementary information can be gained about the effects of background noise on speech processing by breaking down response waveforms into individual oscillatory frequency bands of interest underlying event-related potential components. For instance, our results demonstrate that background noise impacted trial-by-trial neural synchrony during active speech discrimination tasks and that that ITPC and ERSP measures are able to significantly predict speech perception. These results demonstrate that noise-induced disruption of event-related cortical oscillations may be functionally associated with

variability in behavioral speech perception at both the segmental- and sentence-levels in background noise.

Due to time limitation in the current design, this work only used a speech-babble background noise set at one SNR in a double-oddball paradigm. Future research is needed to examine how AERPs and event-related cortical oscillations contribute to the neural generation of the cortical P3 response and how these measures are associated with behavioral performance in various listening conditions and populations with different degrees of language proficiency and degrees of hearing loss. As AERPs are non-invasive and used in difficult-to-test clinical populations for objective assessment of perceptual thresholds (Burkard, Don & Eggermont, 2012), establishing correlations between AERP measures and cognitive measures with a range of speech-in-noise abilities would help elucidate the functional significance of the current electrophysiological data. In addition, the MMN and P3 paradigms can include additional noise types, noise levels, as well as speech and non-speech stimuli and be tested with children, adults, and elderly to study age-related developmental aspects of speech-in-noise processes.

V. Conclusions

This incremental work was designed to evaluate noise-induced effects on the speech-evoked cortical P3 response as well as event-related cortical oscillations in delta, theta, and alpha frequency bands in response to a consonant change and a vowel change in a double-oddball paradigm. Liner-mixed effects models were used to determine whether speech-evoked P3 latency, amplitude, ITPC, and ERSP were predictive of phoneme- and sentence-level speech recognition across listening conditions (quiet vs.

noise), stimuli (/da/ vs. /bu/), and participants. As predicted, the presence of background noise significantly increased P3 latency, decreased P3 amplitude, and decreased ITPC in associated frequency bands without significant noise-induced effects on trial-by-trial power change across frequency bands. Consistent with behavioral performance at the segmental level, background noise had a larger impact on the neural processing of the consonant change in comparison with detecting the vowel change. This pattern of differential noise-induced impacts on consonant vs. vowel discrimination in the P3 response was opposite to our previous MMN study (Koerner et al, 2016), indicating the important role of attention in modulating the ERP components of interest. Moreover, the cortical P3 response and its associated event-related cortical oscillations represent potential neural markers for speech perception at both segmental and sentence-levels. This work has important implications regarding the clinical utility of the P3 response that should be validated in future studies using different stimuli, listening conditions, and participant populations.

Chapter 6: The Role of Hearing Loss in the Sensory and Cognitive Processing of Speech in Noise

I. Introduction

Listeners with hearing impairment (HI) often show a wide range of performance on speech in noise tasks, which can be systematically examined with non-invasive electrophysiological measures to determine the timing and magnitude of neural responses to speech along the auditory pathway. Studies have established that the presence of background noise can lead to a significant delay and reduction in auditory event-related potentials (AERPs) to speech as well as non-speech stimuli (Billings, Tremblay, Stecker, & Tolin, 2009; Koerner & Zhang, 2015; Kozou et al., 2005; Muller-Gass, Marcoux, Logan, & Campbell, 2001; Whiting, Martin, & Stapells, 1998). Furthermore, previous work has shown that noise-induced variability in AERPs is correlated with changes in the ability to perceive speech in background noise (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2013; Anderson, Parbery-Clark, Yi, & Kraus, 2011; Bennett, Billings, Molis, & Leek, 2012; Billings, Mcmillan, Penman, & Gille, 2013; Koerner, Zhang, Nelson, Wang, & Zou, 2016; Song, Skoe, Banai, & Kraus, 2011). But the previous studies did not examine listeners with various degrees of hearing loss, and it remains unclear whether electrophysiological measures can be a reliable predictor of speech perception performance in noise in individuals with HI. The present ERP study aimed to examine the role of hearing impairment in the neural coding of speech sounds in terms of brain-behavior correlates in this clinical population in order to better understand the neural mechanisms underlying speech perception in these individuals.

Previous studies have examined the effects of reduced audibility on speech perception using N1, MMN, N2, and P3 AERPs in response to a /ba/-/da/ stimulus contrast (Martin, Sigal, Kurtzberg, & Stapells, 1997; Martin & Stapells, 2005; Martin, Kurtzberg, & Stapells, 1999). To have a better control of subject characteristics, these studies tested noise-masking conditions in normal hearing listeners in order to simulate reduced audibility caused by different degrees and configurations of hearing loss. The results showed that when audibility was reduced, response latencies increased and amplitudes decreased. In addition, reduced audibility from noise masking impacted behavioral discrimination and reaction time in response to the /ba/-/da/ speech contrast. The effects of reduced audibility were largest when noise masking encompassed the 1-2 kHz frequency range, which contains acoustic cues that are important for accurate discrimination of the /ba/-/da/ stimulus contrast. However, there was a differential effect of noise on these responses, such that N1 was present as long as stimuli were audible, while later AERP components such as the P3 and MMN were present only if stimuli were discriminable. Thus, while each AERP component appeared to be sensitive to the effects of reduced audibility on speech perception, results showed that later cortical responses tended to be more susceptible to the effects of noise masking.

Elicited AERPs from individuals with HI using speech stimuli in quiet have been mixed. Early work showed that N1, P2, N2, and P3 AERPs are significantly delayed in participants with HI compared to normal hearing (NH) control participants. However, while P2 and N2 AERP peaks were significantly smaller in participants with HI, the amplitudes of N1 and P3 peaks were not significantly different from NH controls (Polen,

1984). In contrast, a study that examined N1, P2, and P3 AERPs in participants with HI showed that N1 amplitude was the only metric that differed significantly from NH control participants (Wall, Dalebout, Davidson, & Allen Fox, 1991). In an attempt to examine the effects of different degrees of HI on AERPs as well as behavioral measures of speech discrimination, Oates, Kurtzberg, and Stapells (2002) found that HI caused N1, MMN, N2, and P3 amplitude decreases and latency increases in response to speech stimuli in quiet. Similarly, behavioral measures of phoneme discrimination sensitivity were significantly impacted. As expected, it was also shown that and these hearing-related effects on the neural coding and perception of speech become larger with greater decreases in hearing sensitivity. Moreover, similar to results from earlier noise-masking studies (Martin et al., 1997; Martin & Stapells, 2005; Martin et al., 1999), effects of HI appeared to be larger for later AERP components, such as the P3 response, compared to earlier N1 and MMN responses (Oates, Kurtzberg, & Stapells, 2002). Differences in results across these studies may be caused by dissimilarities in experimental methodology, including speech stimuli and presentation levels, as well as variability in participant groups. A large barrier to examining individuals with HI is the inherent lack of homogeneity across participants, which also causes difficulty in the interpretation of results. For example, even if participants are grouped based on audiometric hearing thresholds, there may be large differences in the etiology and length of hearing impairment as well as suprathreshold auditory processing abilities, which could greatly impact the neural coding and perception of speech. Furthermore, it is difficult to control for other confounds such as the effects of aging on auditory processing as well as

unknown effects due to other co-occurring health conditions. In any case, results from these previous studies suggest that reduced audibility from HI impacts the neural coding of speech in quiet, which tends to manifest as increases in AERP latencies and decrease in AERP amplitudes. Moreover, evidence suggests that HI may differentially impact the neural coding of speech at different stages in auditory processing.

Much less is known about how HI impacts the cortical encoding of speech in background noise. Billings et al. (2015) represents the first work to examine AERPs in response to speech in noise in participants with HI. Although there was a significant effect of HI on behavioral measures of speech perception, their results showed that HI did not significantly impact cortical speech-evoked P1-N1-P2 responses in noise. In order to further examine the effects of HI on different levels of cortical processing in response to speech in noise, the current study examined N1-P2 and MMN responses in adult listeners with and without HI. Cortical event-related potentials components can be divided into exogenous, sensory potentials, or endogenous, cognitive potentials. The N1-P2 complex is an obligatory response that is thought to reflect the sensory encoding of audible, acoustic information at the auditory cortex. On the other hand, the mismatch negativity (MMN) reflects pre-attentive perceptual and cognitive processing mechanisms responsible for sensory discrimination. This contrast is evident in results from Martin and colleagues (Martin et al., 1997; Martin & Stapells, 2005; Martin et al., 1999), who showed that the N1 response was present as long as speech stimuli were audible, regardless of whether they were discriminable, while the MMN response became absent as soon as participants were no longer able to discriminate speech stimuli. Similarly,

research examining categorical perception using voice-onset time (VOT) contrasts has shown that while the N1 response is influenced by acoustic characteristics of the stimulus, it does not represent a neural correlate of VOT perception (Sharma, Marsh, & Dorman, 2000). In contrast, the strength of the MMN response reflected VOT discrimination, such that MMN responses to distinct across-category stimuli was larger and more robust than MMN responses to within-category stimulus pairs (Sharma & Dorman, 1999). Together, the N1-P2 and MMN AERP components can provide information about the effects of various factors, such as the presence of HI or background noise, on different levels of cortical processing underlying auditory and speech perception. To our knowledge, no studies have directly compared the systematic effects of HI on these cortical responses to speech in noise.

When paired with behavioral speech perception tests, AERPs can be used to examine neural correlates of behavior and to better understand mechanisms underlying impaired performance in background noise. Billings et al. (2015) showed that passive N1 and P2 responses were able to predict the effects of HI on sentence-level perception in background noise. Less is known about whether the passive MMN response is a neural correlate of hearing-related changes in speech perception in noise. It is well known that the MMN has high inter- and intra-subject variability (Kurtzberg et al., 1995; Lang et al., 1995; Martin et al., 1999, 2008; Näätänen 2007; 1995; Stapells, 2002), which greatly limits its use in assessing performance at the individual level. However, recent findings suggest that a measure of cortical oscillatory activity associated with the MMN response may represent a more robust measure of neural processing than averaged MMN latency

or amplitude (Koerner et al., 2016). It has been established that oscillatory rhythms in the theta band are linked with cognitive memory processes and likely contribute to the neural generation of the MMN response (Choi et al., 2013; Fuentemilla et al., 2007; Hsiao et al., 2009; Ko et al., 2012). Traditional AERP waveform averaging focuses on capturing synchronous neural activity that is time-locked to an auditory stimulus, but does not allow for an examination of ongoing EEG oscillatory activity underlying AERP components, as any "random" trial-by-trial activity is canceled out during the averaging process. Trial-by-trial latency jitter induced by factors such as HI or background noise is likely linked to reduced AERP amplitudes and delayed AERP latencies in the averaging process. Therefore, not only are averaged AERP components distorted by this trial-by-trial variation, but they are also unable to quantify useful information from this underlying oscillatory activity. This study was designed to determine whether the MMN and its associated event-related cortical oscillations are able to predict the effects of HI on speech-in-noise perception.

The current work extended a double oddball paradigm used in our previous studies on NH listeners (Koerner et al., 2016; Koerner et al., 2017) to examine AERPs and event-related cortical oscillations in response to two speech contrasts in background noise. Behavioral tests of phoneme discrimination and sentence recognition were included to examine potential brain-behavior relationships. The double oddball paradigm was used to compare responses to a consonant change with weaker acoustic cues to that of a more salient and stable vowel change. This paradigm allows for a within-participant control condition, as responses to the two deviant stimuli that were recorded within the

same testing session can be compared. All AERP and behavioral measures were recorded in two listening conditions: speech-shaped background noise at a 0 dB SNR and at a -3 dB SNR. These SNRs were chosen to systematically reduce audibility across NH and HI participants. Articulation Index represents an estimation of audibility, which can be impacted by reduced hearing thresholds as well as by varying signal and noise levels. A measure of AI was used to ensure that at a presentation level of 70 dB SPL, only about 20-40% (AI values of 0.2 to 0.4) of the speech signal would be audible for participants with hearing thresholds ranging from within normal limits to mild sloping to severe sensorineural hearing losses. An AI of around 0.3 has previously been shown to be a point at which listeners show wide variability in their abilities to perceive speech in noise (Nelson et al., 2012). This AI range was chosen to ensure that participants had a range of performance on behavioral speech recognition tasks without reaching ceiling or floor performance.

This study was designed to examine the effects HI on the neural coding and perception of speech in noise. It was expected that participants with poorer hearing thresholds would have lower performance on behavioral tasks, prolonged AERP latencies, as well as reduced AERP amplitudes and cortical oscillatory activity. However, it was expected that HI would have a differential effect on N1-P2 and MMN responses, such that sensory processing of acoustic speech cues would be less impacted by HI compared to later mechanisms that reflect auditory change discrimination. Because background noise has an effect on audibility, it was predicted that a change in SNR would also have a significant impact on neural and behavioral responses to speech.

In addition, it was predicted that, consistent with previous work, neural and behavioral responses to the consonant change would be weaker than that to the vowel change. It was also expected that objective N1-P2 and MMN responses as well as their associated event-related cortical oscillations would be significantly predictive of phoneme- and sentence-level behavioral performance across listening conditions and participants with various degrees of hearing sensitivity.

II. Materials and Methods

A. Participants

The participants in this study were 11 right-handed, native speakers of American English (age range = 40-71 years, 5 males, 6 females) with hearing sensitivity ranging from within-normal-limits to moderate-to-severe sensorineural hearing loss (Figure 1), as determined by a standard audiological assessment for pure tones from 0.25 – 8 kHz. Participants had no history of speech, language, or cognitive difficulty. The Human Research Protection Program at the University of Minnesota approved the research protocol and al participant provided informed consent prior to beginning the study.

In order to assess the effects of HI on the neural coding and perception of speech in noise, a two-frequency pure-tone average (PTA^2) was calculated for each participant based on the average of hearing thresholds at 1 and 2 kHz in the better ear. Similar PTA measures have been used in previous studies that examined the effects of reduced audibility or HI on neural responses to CV speech syllables (Oates et al., 2002); however, instead of using this measure to group participants into categories that quantify the degree of hearing loss (i.e. “mild” or “severe”), the current study used PTA^2 as a continuous

variable. This method better quantifies variability across participants that would normally be grouped together, which allowed for a better examination of how variability in hearing thresholds across participants impacts the neural coding of speech.

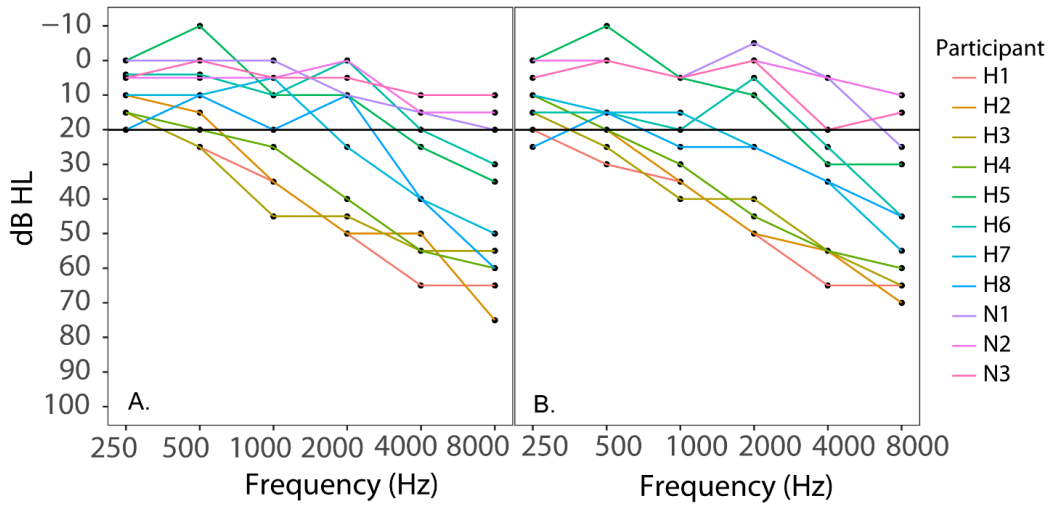

Figure 10. Pure-tone hearing thresholds from .25 to 8 kHz for the left (A) and right (B) ears of each participant.

B. Stimuli

A passive double-oddball paradigm was developed using the consonant-vowel (CV) syllables, /ba/, /da/, and /bu/, to elicit AERP responses. Detailed methodology regarding the creation of these stimuli as well as specific formant frequencies for the formant transitions and steady vowel portions of each CV syllable have been described previously (Koerner et al., 2016). These CV syllables were also used in an active listening condition to obtain phoneme-change detection sensitivity and reaction time. Participants were also presented with IEEE sentences (IEEE, 1969) to obtain sentence-level recognition scores.

Speech and noise stimuli for the EEG test sessions were presented using EEvoke software (ANT Inc., Netherlands) and Goldwave (Goldwave, Inc., 2015) while all behavioral test materials were presented using MATLAB. All CV and sentence-level speech stimuli were presented in speech-shaped background noise that was created using the long-term speech spectrum of the CV syllables and the IEEE sentence corpus, respectively. The root mean square (rms) value for each speech and noise stimulus was normalized so that speech stimuli were always presented at 70 dB SPL in 0 and -3 dB signal-to-noise ratio (SNR) listening conditions. All stimuli were presented via two sound field speakers (M-audio BX8a) located at 45 degrees to the left and right of the participant at a distance of approximately 1 meter.

C. Procedure

Testing was completed in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems). The reported AERP data were taken from a larger scale two-hour EEG recording session that included both passive and active listening conditions. Details regarding stimulus presentation have been reported previously (Koerner et al., 2016). Stimulus presentation order was pseudo-randomized in a double-oddball paradigm so that no blocks began with a deviant stimulus and so that two deviants were never presented in succession. The double-oddball paradigm contained two speech contrasts: a vowel change (from /ba/ to /bu/) and a consonant change (from /ba/ to /da/). The standard /ba/ stimulus had a probability of occurrence of 0.75 and a total of 832 trials while the two deviant stimuli, /bu/ and /da/, each had a probability of occurrence of 0.125 and a total of 104 trials for each stimulus. The SNR listening

conditions were counter-balanced across participants to avoid potential order effects. Each condition consisted of 10 blocks with a 5 second inter-block interval. The interstimulus interval (ISI) was randomized from trial to trial within the range of 600-700 ms. Both deviant stimuli were presented as standard stimuli in 4 separate alternating blocks of 30 repetitions each, for a total of 120 repetitions of each stimulus. These "standard" presentations of /bu/ and /da/ elicited the N1-P2 responses and were subtracted from the deviant stimuli in the double-oddball paradigm to obtain MMN responses. This resulted in an "identity MMN" for each deviant stimulus, which avoids acoustic confounds between the standard and deviant stimuli in interpreting the MMN data (Kraus et al., 1995; 1995; Zhang et al., 2005; Pulvermuller and Shtyrov, 2006). Participants were instructed to relax, minimize excessive movements or eye blinks, and stay awake. During the passive recording session, participants were instructed to ignore the auditory stimuli played through the speakers while focusing on a muted movie of their choice with subtitles.

Behavioral responses for syllable detection were recorded in a separate discrimination task using the same stimuli and presentation levels as in the double-oddball paradigm described above. The participants were asked to focus their attention on detecting phoneme-changes and to make button-press responses using a desktop keyboard whenever they heard a consonant or vowel change. In an additional behavioral testing session, participants listened to randomized lists of IEEE sentences and were instructed to type word by word, as best as they could, the sentences that they heard.

D. Data Analysis

AERP Measures

Continuous EEG data were recorded using the Advanced Neuro Technology EEG System (Advanced Source Analysis version 4.7) and a 64-channel Ag AgCl electrode WaveGuard cap with a REFA-72 amplifier (TMS International BV) (bandwidth = 0.016-200 Hz, sampling rate = 512 Hz). The average impedance of electrodes was below 5 kOhms. ERP waveform analysis was completed offline in BESA (Version 6.0, MEGIS Software GmbH, Germany). The EEG data were bandpassed at 0.5-40 Hz. The ERP epoch length consisted of a 100 ms prestimulus baseline and a 700 ms poststimulus interval. Automatic artifact rejection criteria were set at ± 50 μV. The N1, P2, and MMN responses were analyzed with a common average reference at the Cz electrode. Based on the grand average waveforms in the quiet and noise conditions, N1 and P2 peak latency, relative to the pre-stimulus interval, was measured within a time window of 80-180 ms and 180-300 ms for the 0 dB SNR condition, respectively, and 100-200 ms and 200-350 ms for the -3 dB SNR condition. Similarly, MMN latency was assessed within the time window of 100 -300 ms for the 0 dB SNR condition and 150-350 ms for the -3 dB SNR condition. Amplitude for the N1 and P2 responses were calculated using an averaging window of 10 ms centered at the peak, while MMN amplitude quantification used an integration window of 20 ms centered at peak.

Trial-by-trial phase locking associated with the N1-P2 responses was calculated in delta (0.5-4 Hz), theta (4-8 Hz), and alpha (8-12 Hz) frequency bands using the inter-trial phase coherence (ITPC) measure from the EEGLAB software (Delorme & Makeig, 2004). Inter-trial phase coherence estimates EEG trial-by-trial mean normalized phase as

a function of time and frequency. These values range from 0, which indicates no synchronization across trials, to 1 which indicates perfect synchronization across trials. Inter-trial phase coherence data was averaged across each frequency band at electrode Cz and averaged peak ITPC values in time windows corresponding to the N1 and P2 components was extracted for each frequency band, listening condition, and participant. Additionally, logarithmic spectral power in the theta band was extracted using the subtracted MMN waveform at electrode Cz over the entire analysis epoch, including the pre-stimulus baseline from -100 to 700 ms, using the EEGLAB software (Delorme & Makeig, 2004). These calculations used a modified short-term Fourier Transform (STFT) with Hanning window tapering (Koerner & Zhang, 2015), which is recommended for the analysis of low frequency activities. The modified STFT method in EEGLAB used overlapping sliding windows that are adapted to the target frequency bins to overcome limitations due to the use of fixed windows in conventional analysis. Zero-padding was applied to short epochs with insufficient sample points for Fourier transform. The frequency range for calculating ITPC and spectral power was 0.5 to 40 Hz with a step interval of 0.5 Hz.

All statistical analyses were completed in R (R Core Team, 2014). Linear mixed-effects (LME) models were created, with $\alpha = 0.05$, to examine the statistical significance of stimulus type (/da/ or /bu/), SNR condition (0 or -3 dB SNR), HI (PTA^2) and any potential interactions on MMN latency, amplitude, and spectral power recorded at electrode Cz across all participants.

Behavioral Measures

In the phoneme discrimination test, percent correct scores and reaction time for the detection of consonant and vowel changes in the double-oddball paradigm were obtained from the button-press responses recorded during the quiet and noise conditions. Linear mixed-effects regression models were used to determine statistical significance of stimulus type (/da/ or /bu/), SNR condition (0 or -3 dB SNR), and HI (PTA^2) on behavioral discrimination accuracy and reaction time across all participants.

Participants completed two lists of IEEE sentences (IEEE, 1969) spoken by a female and two lists spoken by a male for each listening condition. Each list consists of 10 low context sentences with 5 key words in each sentence. Therefore, each participant was presented with 4 sentence lists for 200 key words total per listening SNR listening condition. Word-by-word responses were automatically scored by MATLAB and all incorrect responses were checked by a researcher to ensure that spelling mistakes did not result in an incorrect response. An additional LME model was carried out to examine the significance of listening SNR condition (0 or -3 dB SNR) and HI (PTA^2) on sentence recognition across participants.

Brain-Behavior Relationships

Linear mixed-effects (LME) models were developed in R (R Core Team, 2014) and the *nlme* package (Pinheiro et al., 2016) to determine whether objective AERPs and measures of event-related cortical activity were predictive of behavioral speech perception at the syllable- and sentence-levels (Koerner & Zhang, 2017). Data transformations included re-scaling AERP latency and behavioral reaction time values. Participants were used as a "by-subject" random effect in each LME model while speech

stimulus (/bu/ or /da/), SNR condition (0 or -3 dB SNR), and PTA^2 were included as blocking variables to account for repeated measure and inherent differences within and across participants. AERP latency, amplitude, and ITPC or spectral power values were added as fixed effects in order to predict percent correct phoneme detection, reaction time, and sentence recognition scores across participants. Separate models were developed for neural responses to the consonant and vowel change for prediction of sentence-level performance. The significance of each fixed-effect in predicting each behavioral outcome measure was assessed with a $\alpha = 0.05$.

III. Results

Analysis revealed that HI had a significant effect on behavioral speech tasks as well as the MMN response, but did not significantly impact N1 or P2 AERPs (Table 16, Table 17). Results also showed significant effects of stimulus (/bu/ vs. /da/) and SNR listening condition (0 vs. -3 dB SNR) across AERP and behavioral responses (Table 16, Table 17). Linear mixed-effects regression analysis showed that AERPs represent neural correlates of phoneme-and sentence-level performance across participants (Table 18, Table 19, Table 20).

A. Brain measures

Results from the LME models showed that there was a significant effect of stimulus on N1 latency ($F(1,31) = 12.61$, $p < 0.01$). N1 latency was significantly delayed in response to the consonant change compared to the vowel change across listening conditions (Table 16). There was also a significant effect of stimulus on ITPC associated with the P2 response in the theta ($F(1,36) = 8.11$, $p < 0.01$) and alpha ($F(1,36) = 8.96$, $p <$

0.01) frequency bands across listening conditions. Results showed that ITPC in theta and alpha bands was significantly lower in response to the consonant change compared to the vowel change across listening conditions. The LME models also revealed that there was a significant effect of SNR listening condition on N1 amplitude ($F(1,31) = 10.99$, $p < 0.01$) as well as ITPC associated with the N1 response in the delta ($F(1,31) = 9.79$, $p < 0.01$), theta ($F(1,31) = 11.54$, $p < 0.01$), and alpha ($F(1,31) = 6.64$, $p < 0.01$) frequency bands across stimuli. Results showed that N1 amplitude as well as ITPC across frequency bands was smaller in the 0 dB SNR condition compared to the more difficult -3 dB SNR condition. There was also a significant effect of SNR condition on P2 latency ($F(1,31) = 15.11$, $p < 0.001$), such that latency was longer in response to speech stimuli in the -3 dB SNR condition compared to the 0 dB SNR condition.

The linear-mixed effects regression models revealed that there was a significant effect stimulus ($F(1,31) = 5.00$, $p < 0.05$) and listening condition ($F(1,31) = 5.72$, $p < 0.05$) on MMN latency. In general, MMN latency was longer in response to the consonant change compared to the vowel change across conditions and was also longer in the -3 dB SNR condition compared to the 0 dB SNR condition across stimuli. The LME models also revealed a significant effect of hearing loss, as measured by PTA^2, on MMN amplitude ($F(1,9) = 14.01$, $p < 0.01$). Results showed that MMN amplitude was significantly smaller with poorer PTA^2 values across listening conditions and stimuli.

Table 16. F-statistics for effects of stimulus, SNR listening condition, PTA, and resulting interactions across participants for N1, P2, and MMN measures.

N1	Effect of Stimulus	Effect of SNR Condition	Effect of PTA^2
Latency	**12.61****	4.09	0.11
Amplitude	1.53	**10.99****	0.92
Delta ITPC	1.77	**9.79****	0.47
Theta ITPC	3.65.	**11.54****	0.10
Alpha ITPC	1.71	**6.64***	0.07
P2			
Latency	3.88	**15.11*****	1.37
Amplitude	2.36	1.00	0.08
Delta ITPC	2.67	1.35	0.07
Theta ITPC	**8.11****	3.06.	0.60
Alpha ITPC	**8.96****	1.96	0.34
MMN			
Latency	**5.00***	**5.72***	0.51
Amplitude	0.12	0.01	**14.01****
Theta Power	1.90	0.50	1.97

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

B. Behavioral Measures

Results from the LME models showed that there was a significant effect of stimulus (/bu/ vs. /da/) on phoneme change-detection sensitivity ($F(1,27) = 909.76$, $p < 0.001$) and reaction time ($F(1,27) = 7.06$, $p < 0.05$), such that the ability to accurately detect the consonant change was drastically poorer than the ability to detect the vowel change across listening conditions (Table 17). Linear mixed-effects regression analysis also revealed a significant effect of listening condition on percent correct sentence recognition ($F(1,9) = 64.64$, $p < 0.001$), such that performance was poorer in the -3 dB SNR listening condition compared to the 0 dB SNR listening condition across participants. There was also a significant effect of hearing impairment on phoneme-change detection sensitivity ($F(1,31) = 6.06$, $p < 0.05$) and sentence recognition ($F(1,9) =$

6.17, $p < 0.05$) across listening conditions, such that those with a better hearing thresholds had higher behavioral performance scores compared to those with poorer hearing sensitivity.

Table 17. F-statistics for effects of stimulus, SNR listening condition, PTA, and resulting interactions across participants for each behavioral measure.

Behavioral Measure	Effect of Stimulus	Effect of SNR Condition	Effect of PTA
Phoneme Discrimination Accuracy (d')	**909.76*****	3.58	**6.06***
Phoneme Discrimination Reaction Time (ms)	**7.06***	0.01	4.65
Sentence Recognition (% correct)	-	**64.64*****	**6.17***

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

C. Brain-Behavior Relationships

The linear mixed-effects regression models revealed that ITPC in the alpha frequency band associated with both N1 ($F(1,26) = 5.01$, $p < 0.05$) and P2 ($F(1,26) = 4.47$, $p < 0.05$) responses was a significant predictor of phoneme detection reaction time across stimuli, listening conditions, and participants. The LME models used to examine percent correct sentence recognition revealed that delta ITPC associated with the N1 response in response to /da/ ($F(1,26) = 9.40$, $p < 0.05$) as well as P2 latency ($F(1,26) = 8.19$, $p < 0.05$), amplitude ($F(1,26) = 13.27$, $p < 0.05$), and alpha ITPC associated with the P3 response to /da/ ($F(1,26) = 7.31$, $p < 0.05$) were significant predictors of performance (Table 18, 19). The LME models also revealed that theta power was a significant predictor of phoneme change detection reaction time ($F(1,26) = 11.52$, $p < 0.01$) across listening conditions. Additionally, MMN latency in response to the vowel

change (/bu/) was a significant predictor of sentence level perception across listening conditions ($F(1,26) = 7.43$, $p < 0.05$) (Table 20). An examination of residual plots revealed that all residuals were normally distributed without any trend toward heteroscedastic variance for each model.

Table 18. F-statistics for fixed effects (N1 latency, amplitude, and ITPC in delta, theta, and alpha bands) in response to the vowel change (/bu/) and consonant change (/da/) for predicting behavior across participants.

Variable	Percent Correct Phoneme Detection	Phoneme Detection Reaction Time	Percent Correct Sentence Recognition (/bu/)	Percent Correct Sentence Recognition (/da/)
Intercept	839.20***	647.58***	1314.71***	964.29***
Stimulus	869.26***	7.91**	-	-
Condition	3.42	0.01	64.28***	65.32***
PTA	3.99	3.54	14.95**	10.75***
N1 Latency	0.18	0.43	5.53	1.62
N1 Amplitude	0.15	0.13	0.95	0.56
N1 Delta	0.74	0.87	5.92	**9.40***
N1 Theta	0.66	0.13	0.69	0.17
N1 Alpha	0.29	**5.01***	4.67	0.02

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 19. F-statistics for fixed effects (P2 latency, amplitude, and ITPC in delta, theta, and alpha bands) in response to the vowel change (/bu/) and consonant change (/da/) for predicting behavior across participants.

Variable	Percent Correct Phoneme Detection	Phoneme Detection Reaction Time	Percent Correct Sentence Recognition (/bu/)	Percent Correct Sentence Recognition (/da/)
Intercept	1042.64***	1029.41***	538.44***	1029.88***
Stimulus	965.49***	7.76**	-	-
Condition	3.80	0.01	52.44***	191.78***
PTA	6.20*	5.62*	6.00*	11.48**
P2 Latency	0.16	4.05	0.72	**8.19***
P2 Amplitude	0.32	0.16	0.21	**13.27***
P2 Delta	0.37	0.31	1.29	3.63
P2 Theta	4.17	0.96	0.56	0.01
P2 Alpha	2.08	**4.47***	0.10	**7.31***

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 20. F-statistics for fixed effects (MMN latency, amplitude, and theta power) in response to the vowel change (/bu/) and consonant change (/da/) for predicting behavior across participants.

Variable	Percent Correct Phoneme Detection	Phoneme Detection Reaction Time	Percent Correct Sentence Recognition (/bu/)	Percent Correct Sentence Recognition (/da/)
Intercept	860.91***	820.35***	903.70***	800.73***
Stimulus	1013.04***	9.51**	-	-
Condition	3.99	0.02	70.13***	38.36***
PTA	5.11*	4.48	10.07*	8.92*
Latency	0.23	1.89	**7.43***	0.73
Amplitude	1.36	0.00	0.78	2.13
Theta Power	3.52	**11.52****	1.34	0.09

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

IV. Discussion

This work aimed to examine the effects of HI, SNR, and stimulus on N1-P2 and MMN AERPs recorded in response to a consonant versus a vowel change. In addition, this study was designed to determine whether AERPs could predict the effects of HI on phoneme- and sentence-level speech perception in noise.

A. Effects of HI on the Neural Coding and Perception of Speech in Noise

Our preliminary results are consistent with those from Billings et al. (2015), who showed that the presence of hearing loss did not have a significant impact on N1-P2 responses to speech across various presentation levels and SNRs but did affect sentence recognition in noise. The current work adds to this body of knowledge by revealing how HI impacts the encoding of speech in noise at different stages of cortical processing. While HI did not have a significant effect on N1-P2 AERPs or MMN latency in response to speech in noise, our results showed that it did significantly impact MMN amplitude as well as behavioral speech discrimination and recognition tasks across conditions and participants. These results suggest that in the context of these studies, later cortical mechanisms that are involved in sensory and cognitive stimulus discrimination processes may be more sensitive to the effects of HI on the neural coding of speech in noise, while earlier responses that reflect the extraction of acoustic features of speech are less affected by variations in hearing sensitivity across participants. This pattern of results is consistent with those from earlier studies which showed that reduced audibility from simulated HI (Martin et al., 1997; Martin & Stapells, 2004; Martin et al., 1999) and the

presence of actual HI (Oates et al., 2002) tended to have a differential effect on AERP components.

Although they only recorded AERPs in quiet, our findings are consistent with Oates et al. (2002), who showed that HI had a significant effect on MMN amplitude, but not MMN latency in response to speech. A possible explanation for these results, which is supported by conclusions from Oates et al. (2002), is that the presence of HI can induce large variability in AERP amplitude while HI related effects on AERP latency are more predictable and stable. It is possible that the presence of hearing loss exacerbates trial-by-trial latency jitter such that during the averaging process, amplitudes are falsely impacted in the final AERP waveform. This potential issue further highlights the use of time-frequency analysis measures to extract and examine cortical oscillatory activity in frequency bands of interest that are typically ignored by traditional AERP averaging techniques.

B. Effects of Stimuli and SNR Listening Conditions on the Neural Coding and Perception of Speech in Noise

It was expected that the neural coding of the consonant change would be weaker than that of the vowel change for N1, P2, and MMN measures, which would be consistent with data from the behavioral phoneme change detection task. This hypothesis was driven by previous neural and behavioral findings which suggest that weaker, more aperiodic consonants are impacted more in background noise that stable, periodic vowel sounds (Cunningham et al., 2002; Russo et al., 2004; Shetake et al., 2011; Song et al., 2011). The current results showed that N1 latency, MMN latency, behavioral

phoneme change detection did follow this trend. However, the effects of stimulus (/bu/ vs. /da/) on ITPC in the theta and alpha bands associated with the P2 response were unexpected, as trial-by-trial neural synchrony was weaker in response to the consonant change compared to the vowel change across listening conditions and participants. A possible explanation for these results are provided by previous work that compared the passive MMN response to the active P3 AERP in response to the same consonant and vowel stimuli (Koerner et al., 2016; 2017).

Previous results showed a differential effect of background noise on MMN amplitude and theta power; the MMN in response to the more salient vowel change was more affected by background noise than responses to the weaker consonant change (Koerner et al., 2016). However, this effect disappeared in the P3 response when participants actively attended and responded during the double-oddball phoneme discrimination task (Koerner et al., 2017). These results imply that there may be a differential effect of attention on obligatory and active neural responses to these stimuli, such that the relative importance of more salient acoustic cues for the neural processing of speech in noise shifts in the absence of attention. It is possible that the ITPC in the theta and alpha bands associated with the exogenous P2 response are also reflecting this phenomenon. However, this trend was not observed in the passive MMN responses analyzed in the current study. This may be due to differences in participant population and experimental methodology, as the current experiment included participants with varying degrees of hearing sensitivity and used a speech shaped background noise at two SNRs instead of a more difficult 4-talker speech babble. Future studies should focus on

examining the potential differential effects of attention on the neural coding of speech contrasts in noise to determine whether results are consistent across participant populations, speech stimuli, and noise conditions.

C. Neural Correlates of the Effects of HI on Speech-in-Noise Perception

Our results are consistent with previous work that has shown that N1 and P2 latency and amplitude are correlated with the effects of HI on the ability to perceive speech in noise (Billings et al., 2015). The current study also provided novel evidence that a measure of trial-by-trial neural synchrony associated with N1 and P2 responses is a potential indicator of behavioral performance across participants with and without HI. Results showed that the ITPC in the alpha band was a consistent predictor of reaction time for phoneme change detection across both the N1 and P2 responses. The LME models also revealed that alpha ITPC in response to the vowel and consonant changes associated with the N1 and P2 AERPs, respectively, is a potential predictor of sentence level performance in noise. Cortical oscillatory rhythms in the alpha band are thought to be associated with top-down processing mechanisms that allow for the inhibition of irrelevant maskers such as background noise (Klimesch, 2012; Straub et al., 2014; Wilsch et al., 2014). Even though the N1-P2 complex reflects the obligatory encoding of acoustic information at the auditory cortex, it is possible that oscillatory activity in the alpha band is already reflecting top-down inhibitory mechanisms at time points corresponding to these cortical potentials. In other words, greater trial-by-trial phase locking in the alpha band associated with the N1-P2 complex may signal the activation these cortical, top-down processes for inhibiting the neural processing of the noise

stimulus, which would promote higher-level phoneme- and sentence-level perception and therefore reflect perceptual abilities in noise.

Previous studies have not examined whether the MMN response is able to predict the effects of HI on speech perception in noise. However, our data is consistent with previous results that show that the MMN represents a potential neural correlate of speech perception in background noise in participants with NH (Koerner et al., 2016). Previous results from Koerner et al. (2016) showed that MMN amplitude, not MMN latency, in response to the vowel change was a significant predictor sentence-level perception in noise for NH listeners. This is in contrast with the current data that showed that MMN latency recorded in response to the vowel change was able to predict sentence-level perception across listening conditions in participants with various degrees of hearing sensitivity. Similarly, Oates et al. (2002) found that AERP latency was a better predictor of behavior in a quiet listening condition than AERP amplitude in participants with varying degrees of HI. A possible explanation for these conflicting results, as discussed above, is related to the fact that HI may induce larger variability in AERP amplitude that is likely unrelated to hearing-related changes in behavioral performance. Therefore, cortical oscillatory activity in various frequency bands may represent a more robust predictor of behavioral performance compared to averaged MMN responses, as time-frequency analysis measures would be able to capture variability in trial-by-trial activity that may be induced by HI. Consistent with Koerner et al. (2016), the current results showed that spectral power in the theta band is a significant predictor of phoneme change detection across listening conditions and participants. Oscillatory activity in the theta

band is thought to be related to cognitive processing mechanisms that support memory encoding, retrieval and maintenance (Klimesch et al., 2008; Ward, 2003). While it has been established that the endogenous, cognitive MMN response represents and index of auditory change detection, the high inter- and intra-subject variability typically seen in this AERP represents a tremendous barrier to using this passive cortical response for clinical assessment of speech perception difficulties (Kurtzberg et al., 1995; Lang et al., 1995; Martin et al., 1999, 2008; Näätänen, 2007; 1995; Stapells, 2002). Our results suggest that examining cortical oscillations through spectral power analysis may represent an additional, complementary tool to examine mechanisms how various factors impact cognitive processes related to the perception of speech in noise.

D. Novelty, Limitations, and Future Directions

Previous work has shown that HI did not have a significant effect on N1 and P2 responses to speech in noise (Billings et al., 2015). While the current work replicated these findings, it was also one of the first to show that hearing sensitivity may have a differential effect on sensory and cognitive processes during the perception of speech in background noise. More specifically, results suggest that variations in hearing sensitivity may not have the same effect on the early sensory encoding of acoustic cues in noise as on speech discrimination processes and speech perception in background noise. These results provide important information regarding the effects of HI on the encoding of speech-in-noise at different stages of cortical processing and suggest that the MMN may be more sensitive measure to track hearing-related changes in the neural coding of speech in background noise. Future research should compare these AERP responses to speech in

noise in individuals who receive amplification from hearing aids, as this would allow for an examination of whether greater accessibility to speech cues alters the differential effect of HI on neural responses that reflect sensory and cognitive processing mechanisms. These findings could influence rehabilitation strategies for individuals who have difficulty communicating in noise, as solutions targeting top-down cognitive processing mechanisms may be result in improved neural coding and perception of speech in noise compared to those that focus on improving the bottom-up neural processing of acoustic cues.

Previous studies that have examined the neural coding and perception of speech in noise have determined that AERPs represent potential neural correlates of speech perception in noise (Bennett et al., 2012; Billings et al., 2013; 2015; Koerner et al., 2016; Koerner et al., 2017). However, less is known about whether these neural responses are able to track variability in speech-in-noise perception caused by HI. Billings et al. (2015) showed that the cortical N1 and P2 responses are correlated with sentence-level perception in noise in participants with HI. The current work is the first to establish that brain-behavior links between the passive MMN response and the ability to perceive speech in noise remain significant in a population of listeners with and without HI. Furthermore, results showed that time-frequency analysis techniques represent additional tools that can be used to assess speech processing across individuals, which has both theoretical and practical implications. Examining cortical oscillations in frequency bands of interest may be valuable in identifying neural mechanisms that are important for

speech perception in background noise as well as better understanding the neural generation of ERP components.

These novel findings contribute evidence that AERPs may be useful in assessing and predicting performance in clinical populations who have difficulty communicating in complex listening environments. However, future studies with larger sample sizes and more diverse participant populations need to be conducted in order to determine the reliability of these measures prior to use in any clinical applications. At this time, the current work only includes data from 11 participants, which included individuals with completely normal hearing thresholds as well as those with mild sloping to severe bilateral sensorineural hearing losses. Future work should include more participants as well as individuals with more diverse degrees of hearing loss and audiometric configurations.

V. Conclusion

This study was designed to determine how hearing impairment impacts N1-P2 and MMN AERPs and their associated event-related cortical oscillations in response to speech in noise. Furthermore, this work aimed to determine whether these neural measures are able to predict phoneme- and sentence-level perception in background noise across participants with and without HI. Consistent with our predictions, HI had a significant impact on the MMN response, but not on the sensory N1-P2 complex. Results were also consistent with our expectations that these objective neural responses represented a potential predictors of speech-in-noise perception across listening conditions. These findings have important clinical implications regarding the use of

electrophysiological measures in assessing and predicting speech perception in clinical populations with hearing impairment.

Chapter 7: General Discussion & Conclusions

Communication in background noise is difficult for many listeners with or without hearing impairment. Auditory event-related potentials and their associated cortical oscillations in various frequency bands represent objective, non-invasive methods to examine the neural coding of speech. When paired with behavioral measures of speech perception, these electrophysiological tools allow for an examination of the effects HI or background noise on neural mechanisms underlying speech perception. The experiments in this dissertation explored the neural coding of speech in noise in participants with and without HI by examining AERPs and event-related cortical oscillations recorded in response to a consonant and vowel change in background noise. This work also aimed to determine whether these electrophysiological responses are potential neural correlates of speech perception in noise across participants. Key findings from each study are highlighted below:

A. Study 1: Inter-trial Coherence and AERPs in Noise

- The introduction of background noise significantly impacted trial-by-trial neural phase locking across frequency bands without a concomitant reduction in power, which is in support of the "phase resetting" hypothesis that suggests that AERPs are generated by stimulus-induced phase resetting of individual oscillations.
- Variability in ITPC was associated with noise-induced changes in N1 and P2 latencies and amplitudes across participants, suggesting that background noise interrupts cortical phase locking at time points corresponding with N1 and P2 responses, which

is likely responsible for averaged AERP latency increases and amplitude decreases in response to speech in noise.

- Results suggest that time-frequency analysis techniques should be considered in addition to traditional AERP averaging methods, as they are able to provide additional information regarding the importance of different frequency bands for the neural coding of speech in noise and the neural generation of AERPs.

B. Study 2: MMN and Speech Perception in Noise

- The introduction of background noise prolonged MMN latencies, reduced MMN amplitudes and spectral power in the theta band, and reduced the ability to perceive speech at both the phoneme- and sentence-levels. Neural and behavioral responses to the consonant change tended to be weaker across quiet and noise listening conditions; however, there was a differential effect of noise on the neural coding of the consonant and vowel change, such that larger noise-induced reductions in MMN amplitude and theta power occurred in response the vowel change compared to the consonant change.
- Spectral power in the theta band was a significant predictor of MMN latency and amplitude in response to speech in noise, which suggests that noise-induced disruptions in neural synchrony during pre-attentive change detection may modulate the averaged MMN response.
- Noise-induced variability in the MMN and theta power recorded in response to phonemic consonant-vowel contrasts were predictive of both phoneme- and sentence-level speech perception.

C. Study 3: LME Models for Examining Brain-Behavior Relationships

- This technical report provided evidence that Pearson correlations are not appropriate for examining multivariate data that contain built-in differences across- and within-participants. Instead, depending on the research question, LME models are more capable of handling the complex relationships among repeated measures data for exploring brain-behavior relationships.

D. Study 4: P3 and Speech Perception in Noise

- The introduction of background noise significantly prolonged P3 latency, reduced P3 amplitude, and reduced ITPC in delta, theta, and alpha frequency bands.
- Consistent with behavioral results, background noise had a larger effect on P3 responses to the consonant change compared to the vowel change, which, when compared with previous MMN results from study 2, suggests that attention might modulate the relative importance of consonant and vowels stimuli during speech perception in background noise.
- The P3 response as well as measures of trial-by-trial phase locking and power are potential neural correlates for phoneme- and sentence-level speech perception in noise.

E. Study 5: HI and the Neural Coding of Speech in Noise

- Hearing impairment had a significant impact on MMN amplitude in response to the consonant and vowel changes in noise, but did not significantly affect N1-P2 latencies and amplitudes or MMN latencies across participants, which suggests that HI may have a differential effect on the neural coding of speech in noise at different levels of cortical processing.

- Results showed that the N1, P2, and MMN AERPs as well as their associated event-related cortical oscillations represented potential neural correlates of the effects of HI on phoneme- and sentence-level perception in noise.

I. Future Directions & Clinical Implications

The results from this dissertation project established that AERPs and their associated event-related cortical oscillations are potential neural correlates of speech perception in noise across participants with and without HI. Future work will need to further investigate whether these electrophysiological measures are able to reliably predict performance across test sessions and participants in response to different stimuli and listening conditions. Because electrophysiological measures are non-invasive, have fine temporal precision, and are relatively low cost, they represent potential tools to assess and predict the ability to perceive speech in noise in difficult-to-test clinical populations. For instance, neural measures may be used to assess how hearing aid amplification potentially alters the neural coding of speech sounds at different stages of cortical processing (Miller & Zhang, 2014) and whether these hearing aid-related changes in AERPs are linked to improved speech perception, which may inform future development of effective hearing aid algorithms. In addition, AERPs may be useful in assessing whether individuals who cannot behaviorally respond to auditory stimuli, such as infants or developmentally delayed adults, are processing certain speech cues post-hearing aid fitting. For instance, improved N1-P2 responses post-hearing aid fitting could indicate a greater presence of audible acoustic information at the auditory cortex, while improved MMN responses may reflect better pre-attentive speech sound

discrimination with hearing aid amplification compared to an unaided listening condition. A lack of significant hearing aid-related changes in AERPs may inform clinicians about the best course of action for patient rehabilitation. For example, if hearing aid amplification improves audibility but MMN latencies remain delayed, auditory training that focuses on improving auditory processing speed or auditory memory and attention may be recommended (Korczak, Kurtzberg, & Stapells, 2005).

Electrophysiological measure may also be beneficial in examining the time course of perceptual learning or auditory rehabilitation. Auditory training programs are designed to exploit brain plasticity in order to improve speech perception in complex listening situations. Brain imaging tools can be useful in tracking these neurophysiological changes induced by perceptual learning, including measures of neural activation, oscillations, and functional connectivity patterns in the neural substrate dedicated to speech processing (Miller, Zhang, & Nelson, 2016; Rao et al., 2017; Song, Skoe, Banai, & Kraus, 2012; Yu et al., 2017; Zhang & Wang, 2010). Additionally, electrophysiological measures may provide useful information for the development of effective auditory training strategies, as they could track improvements in sensory or cognitive processes underlying speech perception in background noise. For instance, neural measures could be implemented as pre- and post-tests to determine the success of a specific aspect of an auditory training program. Even if the program has yet to result in improved behavioral performance, electrophysiological measures may detect improvements in sensory or cognitive processing mechanisms underlying speech perception, which would support the continued use of the program. In contrast, it could

be determined that training should cease if no measurable improvements in cortical processing are documented.

This work also provided important information regarding the effects of hearing impairment and background noise on the cortical encoding of consonant and vowel cues along the auditory pathway. Theoretical knowledge gained from the current work as well as future studies about the effects of various factors on the neural coding of speech can be used to develop rehabilitation strategies that aim to improve performance in noise. For instance, if it is determined that the relative importance of different auditory cues shift in different listening contexts, hearing aid algorithms or auditory rehabilitation strategies can capitalize on processing certain speech cues that are important for speech perception in different listening environments. Additionally, if it is determined that hearing impairment has a larger impact on cognitive mechanisms underlying speech perception even after auditory cues become more audible from hearing aid amplification, rehabilitation strategies can focus on strengthening neural circuits responsible for cognitive processes such as auditory attention, memory, or processing speed.

Similar to previous studies (Oates et al., 2002; Korzcak et al., 2005), Study 5 estimated individuals' degrees of hearing impairment by using a two-frequency PTA (average of thresholds at 1 and 2 kHz). This measure represents one of many that attempt to estimate peripheral hearing impairment, but it may not be sensitive to suprathreshold spectrotemporal processing abilities that impact auditory perception above and beyond the effects of reduced audibility. In addition, these measure would not be able to identify potential deficits in cognitive mechanisms or central auditory processing abilities. There

is a growing body of evidence showing that a standard pure-tone audiogram is not able to accurately predict difficulties perceiving speech perception in complex auditory environments (Killion & Niquette, 2000; Kujawa & Liberman, 2009), as individuals with NH often report issues communicating in complex auditory environments. For instance, studies examining the effects of blast-related trauma and TBI have shown that cortical AERPs are able to reflect central auditory processing deficits in individuals with normal or near-normal peripheral hearing sensitivity (Gallun et al., 2012). This has important implications regarding the use of AERPs to assess individuals with "hidden hearing loss." Future research should focus on additional measures that may be sensitive to sensory or cognitive processing deficits that are not detected during a typical hearing evaluation. This future work could use procedures similar to those in the current dissertation to examine whether AERPs are able to reflect difficulties perceiving speech in noise in these individuals.

II. Conclusions

The current dissertation work aimed to further understand the effects of hearing impairment and background noise on the neural coding and perception of consonants and vowels and was designed to determine whether electrophysiological measures are predictive of behavioral performance across participants. Findings from each experiment established that event-related cortical oscillations in various frequency bands of interest can provide important, complementary information regarding the neural coding of speech as well as the neural generation of AERPs. In addition, this work provided important information regarding the effects of background noise and hearing impairment on the

neural coding of speech in the auditory cortex. Taken together, findings from Studies 2 and 4 showed that consonants and vowels may differentially contribute to speech perception in noise and that the relative importance of these speech cues shifts depending on the listening context. Findings from Study 5 also showed that hearing impairment may have a differential effect on the neural coding of speech at different stages of cortical processing. While Studies 2 and 4 revealed that AERPs represent neural correlates of speech-in-noise perception in normal hearing participants, Study 5 confirmed that these measures are also predictive of the effects of hearing impairment on behavioral performance. This work adds to our theoretical knowledge regarding the neural coding and perception of speech in noise across individuals with and without HI and has important practical implications regarding the clinical utility of electrophysiological measures for the assessment of speech perception in background noise.

References

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*, 197–207.

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*(5), 970–983.

Anderson, S., Chandrasekaran, B., Yi, H. G., & Kraus, N. (2010). Cortical-evoked potentials reflect speech-in-noise perception in children. *European Journal of Neuroscience*, *32*(8), 1407–1413.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2013). Auditory Brainstem Response to Complex Sounds Predicts Self-Reported Speech-in-Noise Performance. *Journal of Speech, Language, and Hearing Research*, *56*, 31–43.

Anderson, S., Parbery-Clark, A., Yi, H.-G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing*, *32*(6), 750–757.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *The Journal of Neuroscience*, *30*(14), 4922–49226.

Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S., & Kraus, N. (2010). Brainstem correlates of speech-in-noise perception in children. *Hearing Research*, *270*(1–2), 151–157.

Andersson-Roswall, L., Engman, E., Samuelsson, H., & Malmgren, K. (2010). Cognitive outcome 10 years after temporal lobe epilepsy surgery: A prospective controlled study. *Neurology*, *74*(24), 1977–1985.

Ard, M. C., Raghavan, N., & Edland, S. D. (2015). Optimal composite scores for longitudinal clinical trials under the linear mixed effects model. *Pharm Stat*, *14*(5), 418–426.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Speech, Language, Hearing Research*, *41*, 549–563.

Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology*, *37*, 13–20.

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (1999). Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG? *Neuroscience Letters*, *259*, 165–168.

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2000). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, *39*, 241–248.

Başar, E., Demiralp, T., Schürmann, M., Başar-Eroglu, C., & Ademoglu, A. (1999). Oscillatory brain dynamics, wavelet analysis, and cognition. *Brain and Language*, *66*, 146–183.

Basar-Eroglu, C., Basar, E., Demiralp, T., & Schurmann, M. (1992). P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. *International Journal of Psychophysiology*, *13*, 161–179.

Bennett, K. O., Billings, C. J., Molis, M. R., & Leek, M. R. (2012). Neural Encoding and Perception of Speech Signals in Informational Masking. *Ear and Hearing*, *33*, 231–238.

Berger, M. P. F., & Tan, F. E. S. (2004). Robust designs for linear mixed effects models. *Journal of the Royal Statistical Society Series C-Applied Statistics*, *53*, 569–581.

Bidelman, G. M., & Dexter, L. (2015). Bilinguals at the "cocktail party": Dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. *Brain & Language*, *143*, 32–41.

Bidelman, G. M., & Howell, M. (2016). Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception. *NeuroImage*, *124*, 581–590.

Bilgel, M., Prince, J. L., Wong, D. F., Resnick, S. M., & Jedynak, B. M. (2016). A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *Neuroimage*, *134*, 658–670.

Billings, C. J., Bennett, K. O., Molis, M. R., & Leek, M. R. (2011). Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear and Hearing*, *32*(1), 53–60.

Billings, C. J., Mcmillan, G. P., Penman, T. M., & Gille, S. M. (2013). Predicting perception in noise using cortical auditory evoked potentials. *Journal of the Association for Research in Otolaryngology*, *14*, 891–903.

Billings, C. J., Penman, T. M., Mcmillan, G. P., & Ellis, E. M. (2015). Electrophysiology and Perception of Speech in Noise in Older Listeners: Effects of Hearing Impairment and Age. *Ear and Hearing*, *36*, 710–722.

Billings, C. J., Tremblay, K. L., Souza, P. E., & Binns, M. a. (2007). Effects of hearing aid amplification and stimulus intensity on cortical auditory evoked potentials. *Audiology and Neurotology*, *12*(4), 234–246.

Billings, C. J., Tremblay, K. L., Stecker, C., & Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, *254*, 15–24.

Bornkessel-Schlesewsky, I., Philipp, M., Alday, P. M., Kretzschmar, F., Grewe, T., Gumpert, M., … Schlesewsky, M. (2015). Age-related changes in predictive capacity versus internal model adaptability: Electrophysiological evidence that individual differences outweigh effects of age. *Fron Aging Neurosci*, *7*, 217.

Bramhall, N., Ong, B., Ko, J., & Parker, M. (2015). Speech perception ability in noise is correlated with auditory brainstem response Wave I amplitude. *J Am Acad Audiol*, *26*(5), 509–517.

Brunellière, A., Dufour, S., & Nguyen, N. (2011). Regional differences in the listener's phonemic inventory affect semantic processing: A mismatch negativity (MMN) study. *Brain & Language*, *117*, 45–51.

Cahana-Amitay, D., Spiro III, A., Sayers, J. T., Oveis, A. C., Higby, E., Ojo, E. A., … Obler, L. (2016). How older adults use cognition in sentence-final word recognition. *Neuropsychol Dev Cogn B Aging*, *23*(4), 418–444.

Canault, M., Le Normand, M. T., Foudil, S., Loundon, N., & Hung, T. V. (2016). Reliability of the Language Environment Analysis System (LENA (TM)) in European French. *Behavior Research Methods*, *48*(3), 1109–1124.

Caramazza, A., Papagno, C., & Ruml, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language*, *75*, 428–450.

Carreiras, M., & Price, C. J. (2008). Brain activation for consonants and vowels. *Cerebral Cortex*, *18*(July), 1727–1735. https://doi.org/10.1093/cercor/bhm202

Carreiras, M., Vergara, M., & Perea, M. (2009). ERP correlates of transposed-letter priming effects: The role of vowels versus consonants. *Psychophysiology*, *46*, 34–42.

Cheung, M. W. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods*, *13*(3), 182–202.

Choi, J. W., Lee, J. K., Ko, D., Lee, G. T., Jung, K. Y., & Kim, K. H. (2013). Fronto-temporal interactions in the theta-band during auditory deviant processing. *Neuroscience Letters*, *548*, 120–125.

Christmann, C. A., Berti, S., Steinbrink, C., & Lachmann, T. (2014). Differences in sensory processing of German vowels and physically matched non-speech sounds as revealed by the mismatch negativity (MMN) of the human event-related brain potential (ERP). *Brain and Language*, *136*, 8–18.

Cole, R. A., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996). The contribution of consonants versus vowels to word recognition in fluent speech.

Cunningham, J., Nicol, T., King, C., Zecker, S. G., & Kraus, N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hearing Research*, *169*, 97–111.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology*, *112*(5), 758–

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Langauge Research*, *28*(3), 369–382.

Cuthbert, J., P., Pretz, C. R., Bushnik, T., Fraser, R. T., Kolakowsky-Hayner, S. A., Malec, J. F., … Sherer, M. (2015). Ten-year employment patterns of working age individuals after moderate to severe traumatic brain injury: A national institute on disability and rehabilitation research traumatic brain injury model systems study. *Arch Phys Med Rehabil*, *96*(12), 2128–2136.

David, O., Kilner, J. M., & Friston, K. J. (2006). Mechanisms of evoked and induced responses in MEG / EEG. *NeuroImage*, *31*, 1580–1591.

Davidson, D. J., & Martin, A. E. (2013). Modeling accuracy as a function of response time with the generalized linear mixed effects model. *Acta Psychol*, *144*(1), 83–96.

De Kegel, A., Maes, L., Van Waelvelde, H., & Dhooge, I. (2015). Examining the impact of cochlear implantation on the early gross motor development of children with a hearing loss. *Ear Hear*, *36*(3), e113-121.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics (pdf, 0.7 MB). *Journal of Neuroscience Methods*, *134*, 9–21.

Demiralp, T., Ademoglu, A., Istefanopulos, Y., Basar-Eroglu, C., & Basar, E. (2001). Wavelet analysis of oddball P300. *International Journal of Psychophysiology*, *39*, 221–227.

Demiralp, T., Yordanova, J., Kolev, V., Ademoglu, A., Devrim, M., & Samar, V. J. (1999). Time-frequency analysis of single-sweep event-related potentials by means of fast wavelet transform. *Brain and Language*, *66*(1), 129–45.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *PNAS*, *111*(19), 7126–7131.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America*, *76*(1), 87–96.

Eggermont, J. J. (2007). Electric and magnetic fields of synchronous neural activity. In R. F. Burkard, M. Don, & J. J. Eggermont (Eds.), *Auditory Evoked Potentials: Basic Principles and Clinical Application* (pp. 3–21). Philadelphia: Lippincott, Williams, and Williams.

Ergenoglu, T., Demiralp, T., Bayraktaroglu, Z., Ergen, M., Beydagi, H., & Uresin, Y. (2004). Alpha rhythm of the EEG modulates visual detection performance in humans. *Cognitive Brain Research*, *20*, 376–383.

Evans, J., Chu, M. N., Aston, J. A., & Su, C. Y. (2010). Linguistic and human effects on F0 in a tonal dialect of Qiang. *Phonetica*, *61*(1–2), 82–99.

Fogerty, D., & Humes, L. E. (2010). Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors. *Journal of the Acoustical Society of America*, *128*(5), 3114–3125.

Fogerty, D., & Humes, L. E. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *The Journal of the Acoustical Society of America*, *131*(2), 1490.

Fogerty, D., & Kewley-port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *Journal of the Acoustical Society of America*, *126*(2), 847–857.

Fogerty, D., Kewley-Port, D., & Humes, L. E. (2012). The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss. *The Journal of the Acoustical Society of America*, *132*(September), 1667.

Fuentemilla, L., Marco-Pallarés, J., & Grau, C. (2006). Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *NeuroImage*, *30*(3), 909–916.

Fuentemilla, L., Marco-Pallarés, J., Münte, T. F., & Grau, C. (2007). Theta EEG oscillatory activity and auditory change detection. *Brain Research*, *1220,* 93-101.
Gallun, F. J., Diedesch, A. C., Kubli, L. R., Walden, T. C., Folmer, R. L., Lewis, M. S., … Leek, M. R. (2012). Performance on tests of central auditory processing by individuals exposed to high-intensity blasts. *JRRD*, *49*(7), 1005–1024.
Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., … Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, 1–12.
Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., & Olszewski, C. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear Hear*, *28*(3), 412–423.
Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*, 1–42.
Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, *40*, 423–431.
Groenen, P., Snik, A., & van den Broek, P. (1996). On the Clinical Relevance of Mismatch Negativity: Results from Subjects with Normal Hearing and Cochlear Implant Users. *Audiology and Neurotology*, *1*, 112–124.
Gruber, W. R., Klimesch, W., Sauseng, P., & Doppelmayr, M. (2005). Alpha phase syncrhonization predicts P1 and N1 latency and amplitude size. *Cerebral Cortex*, *15*, 371–377.
Haag, N., Roppelt, A., & Heppt, B. (2015). Effects of mathematics items' language demands for language minority students: Do they differ between grades? *Learning and Individual Differences*, *42*, 70–76.
Hadjipantelis, P. Z., Aston, J. A., Muller, H. G., & Evans, J. P. (2015). Unifying amplitude and phase analysis: A compositional data approach to functional multivariate mixed-effects modeling of Mandarin Chinese. *J Am Stat Assoc*, *110*(510), 545–559.
Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., & Bauml, K.-H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage*, *37*, 1465–1473.
Hargus, S. E., & Gordon-Salant, S. (1995). Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment. *Journal of Speech and Hearing Research*, *38*, 234–243.
Hasenstab, K., Sugar, C. A., Telesca, D., McEvoy, K., Jeste, S., & Senturk, D. (2015). Identifying longitudinal trends within EEG experiments. *Biometrics*, *71*(4), 1090–1100.
Hong, L. E., Moran, L. V., Du, X., O'Donnell, P., & Summerfelt, A. (2012). Mismatch negativity and low frequency oscillations in schizophrenia families. *Clinical Neurophysiology*, *123*(10), 1980–1988.

Hsiao, F. J., Wu, Z. A., Ho, L. T., & Lin, Y. Y. (2009). Theta oscillation during auditory change detection: An MEG study. *Biological Psychology*, *81*(1), 58–66.
Hsu, C. H., Lee, C. Y., & Marantz, A. (2011). Effects of visual complexity and sublexical information in the occipitotemporal cortex in the reading of Chinese phonograms: A single-trial analysis with MEG. *Brain and Language*, *117*(1), 1–11.
Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., & Strauser, L. E. (2007). Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effects of presentation rate. *J Speech Lang Hear Res*, *50*(2), 283–303.
Intriligator, J., & Polich, J. (1994). On the relationship between background EEG and the P300 event-related potential. *Biological Psychology*, *37*, 207–218.
Intriligator, J., & Polich, J. (1995). On the relationship between EEG and ERP variability. *International Journal of Psychophysiology*, *20*, 59–74.
Jansson-Verkasalo, E., Korpilahti, P., Jantti, V., Valkama, M., Vainionpa, L., Alku, P., … Näätänen, R. (2004). Neurophysiologic correlates of deficient phonological representations and object naming in prematurely born children. *Clinical Neurophysiology*, *115*, 179–187.
Jin, S.H., & Nelson, P. B. (2006). Speech perception in gated noise: The effects of temporal resolution. *Journal of the Acoustical Society of America*, *119*(5), 3097–3108.
Jin, Y., Diaz, B., Colomer, M., & Sebastian-Galles, N. (2014). Oscillation Encoding of Individual Differences in Speech Perception. *PLOS ONE*, *9*(7), 1–7.
Jouravlev, O., & Lupker, S. J. (2015). Predicting stress patterns in an unpredictable stress language: The use of non-lexical sources of evidence for stress assignment in Russian. *Journal of Cognitive Psychology*, *27*(8), 944–966.
Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., & Muchnik, C. (2006). Identification of syllables in noise: electrophysiological and behavioral correlates. *The Journal of the Acoustical Society of America*, *120*(2), 926–933.
Kaser, M., Soltesz, F., Lawrence, P., Miller, S., Dodds, C., Croft, R., … Nathan, P. J. (2013). Oscillatory Underpinnings of Mismatch Negativity and Their Relationship with Cognitive Function in Patients with Schizophrenia. *PLOS ONE*, *8*(12), 1–11.
Kasisopa, B., Reilly, R. G., Luksaneeyanawin, S., & Burnham, D. (2016). Child readers' eye movements in reading Thai. *Vision Research*, *123*, 8–19.
Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007a). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *122*, 2365–2375.
Killion, M. C., & Niquette, P. A. (2000). What can the pure-tone audiogram tell us about a patient's SNR loss? *The Hearing Journal*, *53*(3), 46–53.
Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617.

Klimesch, W., Freunberger, R., Sauseng, P., & Gruber, W. (2008). A short review of slow phase synchronization and memory: Evidence for control processes in different memory systems? *Brain Research*, *1235*, 31–44.

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews*, *31*(7), 1003–1016.

Ko, D., Kwon, S., Lee, G., Im, H., Kim, H., & Jung, K. (2012). Theta Oscillation Related to the Auditory Discrimination Process in Mismatch Negativity: Oddball versus Control Paradigm, 35–42.

Koerner, T. K., & Zhang, Y. (2015). Effects of background noise on inter-trial phase coherence and auditory N1-P2 responses to speech. *Hearing Research*, *328*, 113–119.

Koerner, T. K., & Zhang, Y. (2017). Application of linear mixed-effects models in human neuroscience research: A comparison with Pearson correlation in two auditory electrophysiology studies. *Brain Sciences*, *7*(26), 1–11.

Koerner, T. K., Zhang, Y., Nelson, P. B., Wang, B., & Zou, H. (2016). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A mismatch negativity study. *Hearing Research*, *339*, 40–49.

Koerner, T.K., Zhang, Y., Nelson, P., Wang, B., & Zou, H. (In press). Neural indices of phonemic discrimination and sentence level speech intelligibility in quiet and noise: A P300 study. *Hearing Research*.

Koerner, T., Zhang, Y., & Nelson, P. (2013). Links between mismatch negativity responses and speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *132*, 2049.

Kolev, V., Demiralp, T., Yordanova, J., Ademoglu, A., & Isoglu-Alkac, U. (1997). Time-frequency analysis reveals multiple functional components during oddball P300. *NeuroReport*, *8*, 2061–2065.

Korczak, P. A., Kurtzberg, D., & Stapells, D. R. (2005). Effects of sensorineural hearing loss and personal hearing aids on cortical event-related potentials and behavioral measures of speech-sound processing. *Ear and Hearing*, *26*, 165–185.

Korczak, P. A., & Stapells, D. R. (2010). Effects of Various Articulatory Features of Speech on Cortical Event-Related Potentials and Behavioral Measures of Speech-Sound Processing. *Ear and Hearing*, *31*, 491–504.

Kozou, H., Kujala, T., Shtyrov, Y., Toppila, E., Starck, J., Alku, P., & Näätänen, R. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research*, *199*(1–2), 31–39.

Kraus, N., & Cheour, M. (2000). Speech Sound Representation in the Brain. *Audiology and Neurotology*, *5*, 140–150.

Kraus, N., Mcgee, T., Carrell, T. D., King, C., Tremblay, K., & Nicol, T. (1995). Central Auditory System Plasticity Associated with Speech Discrimination Training. *Journal of Cognitive Neuroscience, 7,* 25-32.

Kraus, N., McGee, T., Carrell, T. D., & Sharma, A. (1995). Neurophysiologic bases of speech discrimination. *Ear and Hearing*, *16*(1), 19–37.

Kraus, N., Mcgee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory Neurophysiologic Responses and Discrimination Deficits in Children with Learning Problems. *Science*, *273*, 971–973.

Kraus, N., McGee, T., Micco, A., Sharma, A., Carrell, T., & Nicol, T. (1993). Mismatch negativity in school-age children to speech stimuli that are just perceptibly different. *Electroencephalography and Clinical Neurophysiology*, *88*(2), 123–

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the "Critical Period." *Language Learning and Development*, *I*, 237–264.

Kujawa, S. G., & Liberman, M. C. (2009). Adding Insult to Injury: Cochlear Nerve Degeneration after "Temporary" Noise-Induced Hearing Loss. *The Journal of Neuroscience*, *29*(45), 14077–14085.

Kurtzberg, D., Vaughan, H. G., Kreuzer, J. A., & Fliegler, K. Z. (1995). Developmental Studies and Clinical Application of Mismatch Negativity: Problems and Prospects. *Ear and Hearing, 16,* 105-117.

Ladefoged, P. (2006). *A Course in Phonetics* (5th ed.). Boston: Thompson.

Lang, H., Eerola, O., Korpilahti, P., Holopainen, I., Salo, S., & Aaltonen, O. (1995). Practical issues in the clinical application of mismatch negativity. *Ear and Hearing*, *16*(1), 118–130.

Li, N., & Loizou, P. C. (2008). The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise. *Journal of the Acoustical Society of America*, *124*(6), 3947–3958.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185–207.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (2nd ed.). Cambridge, Massachusetts: The MIT Press.

Luger, T. M., Suls, J., & Vander Weg, M. W. (2014). How robust is the association between smoking and depression in adults? A meta-analysis using linear mixed-effects models. *Addictive Behaviors*, *39*(10), 1418–1429.

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, *6*(January), 1–7.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, *86*(4), 283–293.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, *8*(5), 204–210.

Maneshi, M., Moeller, F., Fahoum, F., Gotman, J., & Grova, C. (2012). Resting-state connectivity of the sustained attention network correlates with disease duration in idiopathic generalized epilepsy. *PLoS ONE*, *7*(12), e50359.

Martin, B.A., Kurtzberg, D., & Stapells, D. R. (1999). The Effects of Decreased Audibility Produced by High-Pass Noise Masking on N1 and the Mismatch Negativity to Speech Sounds /ba/ and /da/. *JSLHR, 42*(April), 271–287.

Martin, B.A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds/ba/and/da/. *The Journal of the Acoustical Society of America*, *101*(3), 1585–1599.

Martin, B.A., & Stapells, D. R. (2005). Effects of Low-Pass Noise Masking on Auditory Event-Related Potentials to Speech. *Ear and Hearing*, *26*(2), 195–213.

Martin, B.A., Tremblay, K. L., & Korczak, P. (2008). Speech evoked potentials: From the laboratory to the clinic. *Ear and Hearing*, *29*(3), 285–313.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*(4477), 77–80.

McEvoy, K., Hasenstab, K., Senturk, D., Sanders, A., & Jeste, S. S. (2015). Physiologic artifacts in resting state oscillations in young children: Methodological considerations for noisy data. *Brain Imaging Behav*, *9*(1), 104–114.

Mesgarani, N., David, S. V, Fritz, J. B., & Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *PNAS*, *111*(18), 6792–6797.

Miceli, G., Capasso, R., Benvegnù, B., & Caramazza, A. (2004). The categorical distinction of vowel and consonant representations: evidence from dysgraphia. *Neurocas : Case Studies in Neuropsychology, Neuropsychiatry, and Behavioural Neurology*, *10*(2), 109–121.

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonant. *Journal of the Acoustical Society of America*, *27*(2), 338–352.

Miller, S., & Zhang, Y. (2014). Neural Coding of Phonemic Fricative Contrast With and Without Hearing Aid. *Ear and Hearing*, 1–12.

Miller, S., Zhang, Y., & Nelson, P. (2016). Neural Correlates of Phonetic Learning in Postlingually Deafened Cochlear Implant Listeners. *Ear and Hearing*, 1–15.

Mistridis, P., Krumm, S., Monsch, A. U., Berres, M., & Taylor, K. I. (2015). The 12 years preceding mild cognitive impairment due to Alzheimer's Disease: The temporal emergence of cognitive decline. *J Alzheimers Dis*, *48*(4), 1095–1107.

Muller-Gass, A., Marcoux, A., Logan, J., & Campbell, K. B. (2001). The intensity of masking noise affects the mismatch negativity to speech sounds in human subjects. *Neuroscience Letters*, *299*, 197–200.

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type 1 error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *J Exp Psychol Learn Mem Cogn*, *40*(5), 1287–1306.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear*, *16*(1), 6–18.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., … Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters*, *98*, 217–221.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590.

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, *115*(1), 140–144.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Nash-Kille, A., & Sharma, A. (2014). Inter-trial coherence as a marker of cortical phase synchrony in children with sensorineural hearing loss and auditory neuropathy spectrum disorder fitted with hearing aids and cochlear implants. *Clinical Neurophysiology*, *125*(7), 1459–1470.

Nelson, P., Gregan, M., Nie, Y., Svec, A., Koerner, T., Katare, B. (2012). Variability in speech understanding in noise by listeners with hearing loss. The 164th Meeting of the Acoustical Society of America. Kansas City, Missouri.

Nie, Y., Zhang, Y., & Nelson, P. B. (2014). Auditory stream segregation using bandpass noises: Evidence from event-related potentials. *Frontiers in Neuroscience*, *8*(227), 1–12.

Niemitalo-haapola, E., Haapala, S., Jansson-verkasalo, E., & Kujala, T. (2015). Background Noise Degrades Central Auditory Processing in Toddlers. *Ear and Hearing*, *36*, 342–351.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, *95*(2), 1085–1099.

Niquette, P., Gundmundsen, G., & Killion, M. (2001). QuickSIN Speech-in-Noise Test Version 1.3. Elk Grove Village, IL: Etymotic Research.

Oates, P. a, Kurtzberg, D., & Stapells, D. R. (2002a). Effects of sensorineural hearing loss on cortical event-related potential and behavioral measures of speech-sound processing. *Ear and Hearing*, *23*(5), 399–415.

Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America*, *119*(3), 1727–1739.

Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology*, *82*, 219–226.

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., & Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology*, *118*, 177–185.

Pakarinen, S., Teinonen, T., Shestakova, A., Soo Kwon, M., Kujala, T., Hämäläinen, H., … Huotilainen, M. (2013). Fast parametric evaluation of central speech-sound processing with mismatch negativity (MMN). *International Journal of Psychophysiology*, *87*, 103–110.

Pantev, C., Elbert, T., & Lütkenhöner, B. (Eds.), 1994. *Oscillatory event-related brain dynamics.* Plenum Press, New York.

Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical–cortical relationships tell us about processing speech in noise. *Cognitive Neuroscience*, *33*, 549–557.

Parikh, G., & Loizou, P. C. (2005). The influence of noise on vowel and consonant cues. *Journal of the Acoustical Society of America*, *118*(6), 3874–3888.

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G. M., Mallick, B. K., & Ibrahim, J. G. (2011). A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *Annals of Applied Statistics*, *5*(1), 449–467.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, *52*(11), 1465–1469.

Pedapati, E. V., Gilbert, D. L., Erickson, C. A., Horn, P. S., Shaffer, R. C., Wink, L. K., … Wu, S. W. (2016). Abnormal cortical plasticity in youth with autism spectrum disorder: A transcranial magnetic stimulation case-control pilot study. *J Child Adolesc Psychopharmacol*, *26*(7), 625–631.

Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, *103*(1), 577–587.

Phatak, S. A., & Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *Journal of the Acoustical Society of America*, *121*(4), 2312–2326.

Pickett, J. M. (1957). Perception of Vowels Heard in Noises of Various Spectra. *The Journal of the Acoustical Society of America*, *29*(5), 613–620.

Picou, E. M. (2016). How hearing loss and age affect emotional responses to nonspeech sounds. *J Speech Lang Hear Res*, *59*(5), 1233–1246.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*(4), 465–479.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2016). nlme: Linear and Nonlinear Mixed Effects Models.

Polen, S. B. (1984). Auditory event related potentials. *Seminars in Hearing*, *5*(2), 127–141.

Polich, J. (1997). EEG and ERP assessment of normal aging, *104*(January), 244–256.

Polich, J. (2004). Clinical application of the P300 event-related brain potential. *Physical Medicine and Rehabilitation Clinics of North America*, *15*, 133–161.

Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. S. (2013). Effects of children's working memory capacity and processing speech on their sentence imitation performance. *Int J Lang Commun Disord*, *48*(3), 329–342.

Quene, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425.
R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Radicevic, Z., Vujovic, M., Jelicic, L., & Sovilj, M. (2008). Comparative findings of voice and speech: language processing at an early ontogenetic age in quantitative EEG mapping. *Experimental Brain Research*, *184*, 529–532.
Rao, A., Rishiq, D., Yu, L., Zhang, Y., & Abrams, H. (2017). Neural correlates of selective attention with hearing aid use followed by ReadMyQuips auditory training program. *Ear and Hearing*, *38*(1), 28–41.
Rao, A., Zhang, Y., & Miller, S. (2010). Selective listening of concurrent auditory stimuli: An event-related potential study. *Hearing Research*, *268*(1–2), 123–132.
Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., & Thut, G. (2008). Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. *Cerebral Cortex*, *18*, 2010–2018.
Rong, P., Yunusova, Y., Wang, J., & Green, J. R. (2015). Predicting early bulbar decline in Amyotrophic Lateral Sclerosis: A speech subsystem approach. *Behav Neurol*, *2015*, 1–11.
Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, *115*(9), 2021–2030.
Russo, N., Zecker, S., Trommer, B., Chen, J., & Kraus, N. (2009). Effects of background noise on cortical encoding of speech in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(8), 1185–1196.
Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *62*, 437–448.
Santesso, D. L., Schmidt, L. A., & Trainor, L. J. (2007). Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition*, *65*, 14–21.
Sauseng, P., Klimesch, W., Gerloff, C., & Hummel, F. C. (2009). Spontaneous locally restricted EEG alpha activity determines cortical excitability in the motor cortex. *Neuropsychologia*, *47*, 284–288.
Sauseng, P., Klimesch, W., Gruber, W. R., Hanslmayr, S., Freunberger, R., & Doppelmayr, M. (2007). Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. *Neuroscience*, *146*(4), 1435–1444.
Sharma, A., & Dorman, M., F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, *106*(2), 1078–1083.
Sharma, A., Marsh, C. M., & Dorman, M., F. (2000). Relationship between the N1 evoked potential morphology and the perception of voicing. *Journal of the Acoustical Society of America*, *108*(6), 3030–3035.

Shetake, J. a., Wolf, J. T., Cheung, R. J., Engineer, C. T., Ram, S. K., & Kilgard, M. P. (2011). Cortical activity patterns predict robust speech discrimination ability in noise. *European Journal of Neuroscience*, *34*(11), 1823–1838.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, *12*(4), 283–299.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: Neural correlates. *J Cogn Neurosci*, *23*(9), 2268–2279.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex*, *22*(5), 1180–1190.

Souza, P. E., Boike, K., T., Witherell, K., & Tremblay, K. (2007). Prediction of speech recognition from audibility in older listeners with hearing loss: Effects of age, amplification, and background noise. *Journal of the American Academy of Audiology*, *18*, 54–65.

Spencer, K. M., & Polich, J. (1999). Poststimulus EEG spectral analysis and P300: Attention, task, and probability. *Psychophysiology*, *36*, 220–232.

Spinnato, J., Roubaud, M. C., Burle, B., & Torresani, B. (2015). Detecting single-trial EEG evoked potential using a wavelet domain linear mixed model: Application to error potential classification. *J Neural Eng*, *12*(3), 36013.

Stapells, D. R. (2002). Cortical event-related potentials to auditory stimuli. In J. Katz (Ed.), *Handbook of Clinical Audiology* (6th ed., pp. 395–430). Baltimore: Lippincott, Williams, and Williams.

Straub, A., Wostmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Neuroscience*, *8*, 107.

Stuart, A., & Cobb, K. M. (2015). Reliability of measures in transient evoked otoacoustic emissions with contralateral suppression. *J Commun Disord*, *58*, 35–42.

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*(1), 124–139.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America*, *102*, 3762–3773.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport*, *9*(16), 3557–3560.

Tremblay, K. L., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology*, *114*(7), 1332–1343.

Tsao, F., Liu, H., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Development*, *75*(4), 1067–1084.

van de Velde, M., & Meyer, A. S. (2014). Syntactic flexibility and planning scope: The effect of verb bias on advance planning during sentence recall. *Frontiers in Psychology*, *5*.

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*, 131–156.
Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L., M. (2012). The relationship between high-frequency pure-tone hearing loss, Hearing in Noise Test (HINT) thresholds, and the Articulation Index. *Journal of the American Academy of Audiology*, *23*, 779–788.
Visscher, K. M., Miezin, F. M., Kelly, J. E., Buckner, R. L., Donaldson, D. I., McAvoy, M. P., … Petersen, S. E. (2003). Mixed blocked/event-related designs separate transient and sustained activity in fMRI. *Neuroimage*, *19*(4), 1694–1708.
Wall, L. G., Dalebout, S. D., Davidson, S. A., & Allen Fox, R. (1991). Effect of hearing impairment on event-related potentials for tone and speech discrimination. *Folia Phoniatr*, *43*, 265–274.
Wang, X. F., Yang, Q., Fan, Z., Sun, C. K., & Yue, G. H. (2009). Assessing time-dependent association between scalp EEG and muscle activation: A functional random-effects model approach. *J Neurosci Methods*, *177*(1), 232–240.
Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559.
Whiting, K. A., Martin, B. A., & Stapells, D. R. (1998). The Effects of Broadband Noise Masking on Cortical Event-Related Potentials to Speech Sounds /ba/ and /da/. *Ear and Hearing*, *19*(3), 218–231.
Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., & Obleser, J. (2014). Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cerebral Cortex*, *bhu004*.
Wilson, R. H. (2003). Development of a Speech-in-Multitalker-Babble Paradigm to Assess Word-Recognition Performance. *Journal of the American Academy of Audiology*, *14*, 453–470.
Wilson, R. H. (2011). Clinical experience with the Words-in-Noise Test on 3430 Veterans: Comparisons with pure-tone thresholds and word recognition in quiet. *Journal of the Acoustical Society of America*, *22*, 405–423.
Winkler, I., Kujala, T., Tiitinen, H., Alku, P., Alku, P., Lehtokoski, A., … Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*, 638–642.
Wong, P. C. M., Uppunda, Ajith, K., Parrish, T. B., & Dhar, S. (2008). Cortical Mechanisms of Speech Perception in Noise. *Journal of Speech, Language, and Hearing Research*, *51*, 1026–1041.
Yordanova, J., & Kolev, V. (1998). Event-related alpha oscillations are functionally associated with P300 during information processing. *NeuroReport*, (9), 3159–3164.
Yu, L., Rao, A., Zhang, Y., Burton, P. C., Rishiq, D., & Abrams, H. (2017). Neuromodulatory effects of auditory training on hearing aid use on audiovisual speech perception in elderly individuals. *Frontiers in Aging Neuroscience*, *9*(30), 1–8.

Zenon, A., Klein, P. A., Alamia, A., Boursoit, F., Wilhelm, E., & Duque, J. (2015). Increased reliance on value-based decision processes following motor cortex disruption. *Brain Stimul*, *8*(5), 957–964.

Zhang, Y., Cheng, B., Koerner, T. K., Schlauch, R. S., Tanaka, K., Kawakatsu, M., … Imada, T. (2016). Perceptual temporal asymmetry associated with distinct on and off responses to time-varying sounds with rising versus falling intensity: A magnetoencephalography study. *Brain Sciences*, *6*, 1–25.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., … Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science*, *14*(3), 566–581.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., … Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*, *46*(1), 226–240.

Zhang, Y. & Wang, Y. (2010). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition, 10(2),* 147-160.